

Active Sequential Hypothesis Testing

Mohammad Naghshvar and Tara Javidi

Abstract

Consider a decision maker who is responsible to dynamically collect observations so as to enhance his information in a speedy manner about an underlying phenomena of interest while accounting for the penalty of wrong declaration. The special cases of the problem are shown to be that of noisy dynamic search and variable-length coding with feedback. Due to the sequential nature of the problem, the decision maker relies on his current information state to adaptively select the most “informative” sensing action among the available ones.

In this paper, using results in dynamic programming, lower bounds for the optimal total cost is established. The lower bounds characterize the fundamental limits on the maximum achievable information acquisition rate and the optimal reliability. Moreover, upper bounds are obtained via an analysis of two heuristic policies for dynamic selection of actions. It is shown that the first proposed heuristic achieves asymptotic optimality; where the notion of asymptotic optimality, due to Chernoff, implies that the relative difference between the total cost achieved by the proposed policy and the optimal total cost approaches zero as the penalty of wrong declaration (hence the number of collected samples) increases. Furthermore, using the obtained bounds, the gain of adaptive selection of sensing actions is shown to be at least logarithmic in the penalty associated with wrong declarations. The second heuristic is shown to achieve asymptotic optimality only in a limited setting such as the problem of variable-length coding with feedback. However, by considering the asymptotic where the number of hypotheses is also growing and under a mild technical condition, this second heuristic is shown to achieve non-zero information acquisition rate, establishing a lower bound for the maximum achievable rate. In case of variable-length coding with feedback, this non-zero information rate is shown to be maximum (i.e., any information acquisition at higher rate results in non-zero probability of error), and the proposed heuristic is proved to achieve Burnashev’s optimal error exponent. This result extends the notions of capacity and optimal error exponent to the context of active sequential hypothesis testing.

Index Terms

Active hypothesis testing, sequential analysis, optimal stopping, dynamic programming, feedback gain, optimal error exponent, information acquisition rate.

The material in this paper was presented in part at ISIT 2010, Allerton 2010, Asilomar 2010, ISIT 2011, and CISS 2012.

The authors are with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093 USA. Email: naghshvar@ucsd.edu; tjavidi@ucsd.edu.

This work was partially supported by the industrial sponsors of UCSD Center for Wireless Communication (CWC) and Center for Networked Systems (CNS), and NSF Grants CNS-0533035 and CCF-0729060.

I. INTRODUCTION

This paper considers a generalization of the classical sequential hypothesis testing problem due to Wald [1]. Suppose there are M hypotheses among which only one is true. A Bayesian decision maker is responsible to enhance his information about the correct hypothesis in a speedy and sequential manner while accounting for the penalty of wrong declaration. In contrast to the classical sequential M -ary hypothesis testing problem [2]–[4], our decision maker can choose one of K available actions and hence, exert some control over the collected samples’ “information content.” We refer to this generalization, originally tackled by Chernoff [5], as the *active* sequential hypothesis testing problem.

The active sequential hypothesis testing problem naturally arises in a broad spectrum of applications such as medical diagnosis [6], cognition [7], sensor selections [8], underwater inspection [9], generalized search [10], channel coding with perfect feedback [11]. It is intuitive that at any time instant, an optimized Bayesian decision maker relies on his/her current belief to adaptively select the most “informative” sensing action, i.e., an action that provides the highest amount of “information.” Making this intuition precise is the topic of our study.

The most well known instance of our problem is the case of binary hypothesis testing with passive sensing ($M = 2$, $K = 1$), first studied by Wald [1]. In this instance of the problem, the optimal action at any given time is provided by a sequential probability ratio test (SPRT). There are numerous studies on the generalizations to $M > 2$ ($K = 1$) and the performance of known simple and practical heuristic tests such as MSPRT [2]–[4]. The generalization to the active testing case was considered by Chernoff in [5] where a heuristic randomized policy was proposed and whose asymptotic performance was analyzed. More specifically, under a certain technical assumption on uniformly distinguishable hypotheses, the proposed heuristic policy is shown to achieve asymptotic optimality where the notion of asymptotic optimality [5] denotes the relative tightness of the performance upper bound associated with the proposed policy and the lower bound associated with the optimal policy.

The problem of active hypothesis testing also generalizes another classic problem in the literature: the comparison of experiments first introduced by Blackwell [12]. This is a single-shot version of the active hypothesis testing problem in which the decision maker can choose one of several (usually two) actions/experiments to collect a single observation sample before making the final decision. There have been extensive studies [12]–[18] on comparing the actions. Applying various results from [12], [15] in our context of active hypothesis testing and utilizing a dynamic programming interpretation, an optimal notion of information utility, i.e., an optimal measure to quantify the information gained by different sensing actions, can be derived [19]. Inspired by this view of the problem, which coincides with that

promoted by DeGroot [20], we provide a set of (uniform) lower bounds for optimal information utility. Furthermore, we provide two heuristic policies whose performance is investigated via an asymptotic analysis. The first policy is similar to the one proposed in [19], [21], and is shown to be asymptotically optimal, matching the performance of the scheme proposed in [5] (and follow up works [22], [23]) while relaxing the technical assumption on uniform distinguishability of the hypotheses. In contrast, our second proposed policy is only shown to be asymptotically optimal in the limited setup of binary hypothesis testing, noisy dynamic search, and variable-length coding with feedback. However, this policy has a provable advantage for large M over our first policy, as well as other solutions in the literature. More specifically, this policy can provide, under a mild technical condition, reliability and speedy declaration simultaneously. In information theoretic term, this policy can be shown to achieve non-zero information acquisition rate and hence, to generalize Burnashev's [11] variable-length channel coding scheme. We elaborate on a complete literature survey in Section II-B.

The remainder of this paper is organized as follows. In Section II, we formulate the active sequential hypothesis testing problem and discuss the related works. Section III provides a dynamic programming formulation and characterizes an optimal notion of information utility. In Section IV, we provide three lower bounds and two upper bounds on the optimal information utility. The bounds are complementary for various values of the parameters of the problem. Section V states the consequence of the bounds obtained in Section IV. In particular, the obtained bounds are used to 1) establish notions of order and asymptotic optimality for the proposed policies (in Subsection V-A); 2) characterize lower and upper bounds on the maximum achievable information acquisition rate and the optimal reliability (in Subsection V-B); and 3) derive the advantage of causally and adaptively selecting sensing actions over the best open-loop (randomized) selection rule (in Subsection V-C). In Section VI, we investigate important special cases of the active hypothesis testing including the problems of noisy dynamic search and variable-length coding with feedback. Finally, we conclude the paper and discuss future work in Section VII.

Notation: Let $[x]^+ = \max\{x, 0\}$. A random variable is denoted by an upper case letter (e.g. X) and its realization is denoted by a lower case letter (e.g. x). For any set \mathcal{S} , $|\mathcal{S}|$ denotes the cardinality of \mathcal{S} . For a set \mathcal{A} , let $\Lambda(\mathcal{A})$ denote the collection of all probability distributions on elements of \mathcal{A} , i.e., $\Lambda(\mathcal{A}) = \{\lambda \in [0, 1]^{|\mathcal{A}|} : \sum_{a \in \mathcal{A}} \lambda_a = 1\}$. All logarithms are in base 2. The entropy function on a vector $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_M] \in [0, 1]^M$ is defined as $H(\boldsymbol{\rho}) = \sum_{i=1}^M \rho_i \log(1/\rho_i)$, with the convention that $0 \log \frac{1}{0} = 0$. Finally, the Kullback–Leibler (KL) divergence between two probability density functions $q(\cdot)$ and $q'(\cdot)$ on space \mathcal{Z} is defined as $D(q||q') = \int_{\mathcal{Z}} q(z) \log \frac{q(z)}{q'(z)} dz$, with the convention $0 \log \frac{a}{0} = 0$ and $b \log \frac{b}{0} = \infty$ for $a, b \in [0, 1]$ with $b \neq 0$.

II. PROBLEM SETUP AND SUMMARY OF THE RESULTS

In Subsection II-A, we formulate the problem of active sequential hypothesis testing, referred to as Problem (P) hereafter. Subsection II-B states the main contributions of the paper and provides a summary of related works.

A. Problem Formulation

Here, we provide a precise formulation of our problem.

Problem (P) [Active Sequential Hypothesis Testing]

Let $\Omega = \{1, 2, \dots, M\}$. Let H_i , $i \in \Omega$, denote M hypotheses of interest among which only one holds true. Let θ be the random variable that takes the value $\theta = i$ on the event that H_i is true for $i \in \Omega$. We consider a Bayesian scenario with prior belief $\boldsymbol{\rho}(0) = [\rho_1(0), \rho_2(0), \dots, \rho_M(0)]$, i.e., initially $P(\{\theta = i\}) = \rho_i(0)$ for all $i \in \Omega$. \mathcal{A} is the set of all sensing actions and is assumed to be finite with $|\mathcal{A}| = K < \infty$. \mathcal{Z} is the *observation space*. For all $a \in \mathcal{A}$, the observation kernel $q_i^a(\cdot)$ (on \mathcal{Z}) is the probability density function for observation Z when action a has been taken and H_i is true. We assume that observation kernels $\{q_i^a(\cdot)\}_{i \in \Omega, a \in \mathcal{A}}$ are known and the observations are conditionally independent over time. Let L denote the penalty (loss) for a wrong declaration, i.e., the penalty of selecting H_j , $j \neq i$, when H_i is true.¹ Let τ be the stopping time at which the decision maker retires. The objective is to find a sequence of sensing actions $A(0), A(1), \dots, A(\tau - 1)$, a stopping time τ , and a declaration rule $d: \mathcal{A}^\tau \times \mathcal{Z}^\tau \rightarrow \Omega$ that collectively minimize the total cost $\mathbb{E}[\tau + L\mathbf{1}_{\{d(A^\tau, Z^\tau) \neq \theta\}}]$, where the expectation is taken with respect to the initial belief as well as the distribution of observation sequence. The objective of Problem (P) can be written as to

$$\text{minimize } \mathbb{E}[\tau] + L\text{Pe}, \quad (1)$$

where $\text{Pe} = P(\{d(A^\tau, Z^\tau) \neq \theta\})$ denotes the probability of making a wrong declaration.

B. Overview of the Results and Summary of the Related Works

The first attempt to solve Problem (P) goes back to Chernoff's work on active binary composite hypothesis testing [5]. Chernoff proposed the following scheme to select actions: At each time t , find the most likely true hypothesis and then select an action that can discriminate this hypothesis the best from each and every element in the set of alternative hypothesis. Chernoff showed that as L goes to infinity, the relative difference between the total cost achieved by his proposed scheme and

¹In general, we can define a loss matrix $[L_{ij}]_{i,j \in \Omega}$, where L_{ij} denotes the penalty (loss) of selecting H_j when H_i is true.

the optimal total cost approaches zero; which he termed as *asymptotic optimality*.² One of the main drawbacks of Chernoff's asymptotic optimality notion was his neglecting the complementary role of asymptotic analysis in M . In particular, the notion of asymptotic optimality in L falls short in showing the tension between using (asymptotically large number of) samples to discriminate among a few hypotheses with (asymptotically) high accuracy or a (asymptotically) large number of hypotheses with a lower degree of accuracy. Although the scheme proposed in [5] and its subsequent extensions [22]–[29] are asymptotically optimal in L , their provable information acquisition rate is restricted to zero (potentially unbounded number of samples are used to acquire $\log M$ bits of information). Intuitively, the rate of information acquisition under any given heuristic relates to the ratio between $\log M$ and the expected number of samples: the larger this ratio the faster information is acquired.

As elaborated in Section V-B, to obtain asymptotic characterization of the expected optimal cost in a non-zero information rate regime, it is important to propose schemes which scale optimally with M as well. In his seminal paper [11], Burnashev tackled the primal (constrained) version of Problem (P) in the context of channel coding with feedback (in Section VI-C we explain why channel coding with feedback can be interpreted as a special case of Problem (P)) and provided lower and upper bounds on the expected number of samples (or equivalently channel uses) required to convey one of M uniformly distributed messages with a desired probability of error. The lower bound identified the dominating terms in both number of messages and error probability, hence characterized the optimal reliability function (also known as the error exponent) in addition to the feedback capacity (which was known to coincide with the Shannon capacity [30]). In this paper, we generalize this lower bound to the problem of active sequential hypothesis testing, i.e., Problem (P):

- For all achievable rates, we derive three lower bounds on the expected total cost (1). The bounds hold for all prior beliefs and are complementary for various values of L and M . These bounds are collectively used to generalize the (information theoretic) notions of achievable communication rate [31] and error exponent [32] to the context of active sequential hypothesis testing.
- The first and third lower bounds identify the dominating terms in L and hence are useful in establishing asymptotic optimality of order-1 (due to Chernoff [5]) and order-2 in L . Furthermore, from an information theoretic viewpoint, these bounds are used to characterize an upper bound on reliability function (error exponent) at zero information rate.

²In [5], the objective was to minimize $c\mathbb{E}[\tau] + P_e$ and the proposed policy was shown to be asymptotically optimal as $c \rightarrow 0$. It is straightforward to show that for $L = \frac{1}{c}$, this problem coincides with Problem (P) defined in this paper. However, we have chosen $\mathbb{E}[\tau] + LP_e$ as an objective function for Problem (P) because of its Lagrangian relaxation interpretation of an information acquisition problem in which the objective is to minimize $\mathbb{E}[\tau]$ subject to $P_e \leq \epsilon$ where $\epsilon > 0$ denotes the desired probability of error.

- We use the second lower bound as a converse (in a fashion somewhat similar to the Shannon's channel coding converse [31]) to derive an upper bound \bar{I}_{max} on the maximum achievable information acquisition rate. Additionally, this lower bound allows us to 1) provide an upper bound on the reliability function (error exponent) for all rates $R \in [0, \bar{I}_{max}]$, and 2) establish order optimality in M as a necessary condition for any policy which achieves non-zero information acquisition rate.

In addition to a lower bound on an expected number of samples, Burnashev proposed a coding scheme with two phases of operation whose performance provides a tight upper bound (in both L and M). It is interesting to note that the scheme of Chernoff, if specialized to channel coding with feedback, coincides with the second phase of Burnashev's scheme. However, while the first phase of Burnashev's scheme can achieve any information rate up to the capacity of the channel, Chernoff's one-phase scheme has a rate of information acquisition equal to zero. Inspired by Burnashev's coding scheme, we also obtain two heuristic two-phase policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$:

- Policy $\tilde{\pi}_1$, in its first phase, selects actions in a way that all pairs of hypotheses can be distinguished from each other; while its second phase coincides with Chernoff's scheme [5] where only the pairs including the most likely hypothesis are considered. The second phase of policy $\tilde{\pi}_1$ is shown to ensure its asymptotic optimality in L ; while its first phase in a very natural manner relaxes the technical assumption in [5] where all actions are assumed to discriminate between all hypotheses pairs or the need for the infinitely often reliance on randomized action deployed in [23] in order to ensure sufficient discrimination among hypotheses.
- Policy $\tilde{\pi}_2$, in contrast, is only shown to be asymptotically optimal in three important special cases discussed in Section VI. However, under a mild technical condition, policy $\tilde{\pi}_2$ can ensure that information acquisition occurs at a non-zero rate. Mathematically, this means that, under policy $\tilde{\pi}_2$, the expected total cost (1) grows in L and M in an order optimal fashion establishing a lower bound on the maximum achievable information acquisition rate $\underline{I}_2 \leq \bar{I}_{max}$ as well as a lower bound on the optimal reliability function (optimal error exponent) for all rates $R \in [0, \underline{I}_2]$. As a corollary to the asymptotic optimality result of Section VI, we recover Burnashev's optimal reliability in case of variable-length coding with feedback.

The above results are also used to answer a fundamental question regarding the significance of adaptive decision making. In particular, specializing the obtained lower bounds to the open-loop setup along with our policy $\tilde{\pi}_1$, we investigate the benefit of adapting sensing actions at any decision epoch. We show that almost in all practical settings the adaptivity gain grows logarithmically with penalty L . Such a characterization complements the more recent work on multi-stage policies (introduced in [33], [34])

in which the decision maker can take a retire/declare action only at the end of each stage (of potentially unequal length). Extending our characterization of the adaptivity gain to quantifying the loss in performance introduced by the multi-stage decision making constraint remains an area of future work.

III. DYNAMIC PROGRAMMING AND CHARACTERIZATION OF AN OPTIMAL POLICY

In this section, we first derive the corresponding dynamic programming (DP) equation for Problem (P). From the DP solution, we characterize an optimal policy for Problem (P).

The problem of active M -ary hypothesis testing is a partially observable Markov decision problem (POMDP) where the state is static and observations are noisy. It is known that any POMDP is equivalent to an MDP with a compact yet uncountable state space, for which the belief of the decision maker about the underlying state becomes an information state [35]. In our setup, thus, the information state at time t is nothing but the belief vector $\boldsymbol{\rho}(t)$ whose i^{th} element is the conditional probability of hypothesis H_i to be true given the initial belief and all the observations and actions up to time t . Accordingly, the information state space is defined as $\mathbb{P}(\Theta) := \left\{ \boldsymbol{\rho} \in [0, 1]^M : \sum_{i=1}^M \rho_i = 1 \right\}$ where Θ is the σ -algebra generated by random variable θ . In one sensing step, the evolution of the belief vector follows Bayes' rule and is given by Φ^a , a measurable function from $\mathbb{P}(\Theta) \times \mathcal{Z}$ to $\mathbb{P}(\Theta)$ for all $a \in \mathcal{A}$:

$$\Phi^a(\boldsymbol{\rho}, z) := \left[\rho_1 \frac{q_1^a(z)}{q_\rho^a(z)}, \rho_2 \frac{q_2^a(z)}{q_\rho^a(z)}, \dots, \rho_M \frac{q_M^a(z)}{q_\rho^a(z)} \right], \quad (2)$$

where $q_\rho^a(z) = \sum_{i=1}^M \rho_i q_i^a(z)$, and $\Phi^a(\boldsymbol{\rho}, z) = \boldsymbol{\rho}$ if $q_\rho^a(z) = 0$. In other words, if $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$ is an apriori distribution, $\Phi^a(\boldsymbol{\rho}, z)$ gives us the posteriori distribution when sensing action a has been taken and z has been observed. Note that the posterior distribution is strongly dependent on the sensing action a .

We define operator \mathbb{T}^a , $a \in \mathcal{A}$, such that for any measurable function $g : \mathbb{P}(\Theta) \rightarrow \mathbb{R}$,

$$(\mathbb{T}^a g)(\boldsymbol{\rho}) := \int g(\Phi^a(\boldsymbol{\rho}, z)) q_\rho^a(z) dz. \quad (3)$$

Given that $\boldsymbol{\rho}$ is an apriori distribution and action a has been taken, $(\mathbb{T}^a g)(\boldsymbol{\rho})$ is the expected value of function g at the posterior belief, where the computation of the posterior belief follows Bayes rule as shown in (2). Note that using operator \mathbb{T}^a , one can compute the mutual information between θ with distribution $\boldsymbol{\rho}$ and observation Z under action a with distribution q_ρ^a , i.e.,

$$\begin{aligned} I(\boldsymbol{\rho}; q_\rho^a) &:= H(\boldsymbol{\rho}) - \int H(\Phi^a(\boldsymbol{\rho}, z)) q_\rho^a(z) dz \\ &= H(\boldsymbol{\rho}) - (\mathbb{T}^a H)(\boldsymbol{\rho}). \end{aligned}$$

Fact 1 (Consequence of Propositions 9.8 and 9.10 in [36]). *Let $V^* : \mathbb{P}(\Theta) \rightarrow \mathbb{R}_+$ be the minimal solution to the following fixed point equation:*

$$V^*(\boldsymbol{\rho}) = \min \left\{ 1 + \min_{a \in \mathcal{A}} (\mathbb{T}^a V^*)(\boldsymbol{\rho}), \min_{j \in \Omega} (1 - \rho_j) L \right\}. \quad (4)$$

Then $V^(\boldsymbol{\rho}(0))$, referred to as the optimal value function, is equal to the minimum cost in Problem (P) with the prior belief $\boldsymbol{\rho}(0)$.*

Definition 1. A Markov stationary *policy* is a stochastic kernel from the information state space $\mathbb{P}(\Theta)$ to $\mathcal{A} \cup \{d\}$ describing the conditional distribution on sensing actions $A(t)$, $t = 0, 1, \dots, \tau - 1$ and stopping time τ (the choice of declaration d marks the stopping time τ). In other words, under policy π , the probability that action a is selected at belief state $\boldsymbol{\rho}$ is given by $\pi(a|\boldsymbol{\rho})$.

Definition 2. A policy π is referred to as Markov stationary *deterministic* if for each $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$, there exists an action $a \in \mathcal{A} \cup \{d\}$ for which $\pi(a|\boldsymbol{\rho}) = 1$.

As shown in Corollary 9.12.1 in [36], equation (4) provides a characterization of an optimal Markov stationary deterministic policy π^* for Problem (P) as follows: Sensing action $a^* = \arg \min_{a \in \mathcal{A}} (\mathbb{T}^a V^*)(\boldsymbol{\rho})$ is the least costly sensing action, resulting in $1 + \min_{a \in \mathcal{A}} (\mathbb{T}^a V^*)(\boldsymbol{\rho})$, and is the optimal action to take unless the penalty of wrongly declaring H_{i^*} , where $i^* = \arg \min_{j \in \Omega} (1 - \rho_j) L$, is even less costly in which case it is optimal to retire and declare H_{i^*} as the true hypothesis.

Remark 1. It follows from (4) that if $\min_{j \in \Omega} (1 - \rho_j) L \leq 1$, then $V^*(\boldsymbol{\rho}) = \min_{j \in \Omega} (1 - \rho_j) L$ and hence, the further reduction of the probability of error is not worth taking one more sensing action. Therefore, the region of interest in our analysis is restricted to $L > 1$ and $\mathbb{P}_L(\Theta) := \{\boldsymbol{\rho} \in \mathbb{P}(\Theta) : \min_{j \in \Omega} (1 - \rho_j) L > 1\}$.

Before we close this section, we provide the following lemma.

Lemma 1. *Suppose there exists a functional $V : \mathbb{P}(\Theta) \rightarrow \mathbb{R}_+$ such that for all belief vectors $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$*

$$V(\boldsymbol{\rho}) \leq \min \{ 1 + \min_{a \in \mathcal{A}} (\mathbb{T}^a V)(\boldsymbol{\rho}), \min_{j \in \Omega} (1 - \rho_j) L \}.$$

Then $V(\boldsymbol{\rho}) \leq V^(\boldsymbol{\rho})$ for all $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$ where V^* is a fixed point solution to (4).*

The proof is provided in Appendix I.

IV. PERFORMANCE BOUNDS

As discussed earlier, finding an optimal policy π^* for Problem (P) requires knowledge of the optimal value function V^* . In lieu of numerical approximation of (4) using value iteration techniques [37], or

deriving a closed-form for V^* , in Subsections IV-A and IV-B, we use Lemma 1 and heuristic policies to find lower and upper bounds for the value function V^* respectively.

We have the following technical Assumptions.

Assumption 1. *For any two hypotheses i and j , $i \neq j$, there exists an action a , $a \in \mathcal{A}$, such that $D(q_i^a || q_j^a) > 0$.*

Assumption 2. *There exists $\xi < \infty$ such that $\max_{i,j \in \Omega} \max_{a \in \mathcal{A}} \sup_{z \in \mathcal{Z}} \frac{q_i^a(z)}{q_j^a(z)} \leq \xi$.*

Assumption 1 ensures the possibility of discrimination between any two hypotheses, hence ensuring Problem (P) has a meaningful solution. Assumption 2 implies that no two hypotheses are fully distinguishable using a single observation sample. Assumption 2 is a technical one for ease of our proofs.

A. Lower Bounds for V^*

Proposition 1. *Under Assumption 1 and for $L > 1$, $V^*(\boldsymbol{\rho}) \geq \underline{V}_1(\boldsymbol{\rho})$ where,*

$$\underline{V}_1(\boldsymbol{\rho}) := \left[\sum_{i=1}^M \rho_i \max_{j \neq i} \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i}{\rho_j}}{\max_{a \in \mathcal{A}} D(q_i^a || q_j^a)} - K'_1 \right]^+,$$

and K'_1 is a constant independent of L .

The proof of Proposition 1 is based on Lemma 1 and is provided in Appendix II-A.

Next we provide another lower bound which is more appropriate for large values of M . Remember that $I(\boldsymbol{\rho}; q_\rho^a) = H(\boldsymbol{\rho}) - (\mathbb{T}^a H)(\boldsymbol{\rho})$ denotes the mutual information between $\theta \sim \boldsymbol{\rho}$ and observation Z under action a . Let $D_{max} := \max_{i,j \in \Omega} \max_{a \in \mathcal{A}} D(q_i^a || q_j^a)$, $I_{max} := \max_{a \in \mathcal{A}} \max_{\hat{\boldsymbol{\rho}} \in \mathbb{P}(\Theta)} I(\hat{\boldsymbol{\rho}}; q_{\hat{\boldsymbol{\rho}}}^a)$, and $\alpha(L, M) := \frac{M-1}{M-1+2LI_{max}}$. It can be easily shown that D_{max} and I_{max} are non-decreasing in M and $I_{max} \leq D_{max} \leq \log \xi$ for all M . Moreover, let

$$\begin{aligned} \underline{D}_{max} &:= \inf_M D_{max}, & \overline{D}_{max} &:= \sup_M D_{max}, \\ \underline{I}_{max} &:= \inf_M I_{max}, & \overline{I}_{max} &:= \sup_M I_{max}. \end{aligned}$$

Proposition 2. *Under Assumption 1 and for $L > 1$,*

$$V^*(\boldsymbol{\rho}) \geq \left[\frac{H(\boldsymbol{\rho}) - H([\alpha(L, M), 1 - \alpha(L, M)]) - \alpha(L, M) \log(M-1)}{I_{max}} + \alpha(L, M)L \right]^+.$$

Furthermore, under Assumption 2 and for $L > \frac{\log M}{I_{\max}}$ and arbitrary $\delta \in (0, 0.5]$,

$$V^*(\boldsymbol{\rho}) \geq \underline{V}_2(\boldsymbol{\rho}) := \left[\frac{H(\boldsymbol{\rho}) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{\max}} + \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{\max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - K'_2 \right]^+,$$

where K'_2 is a constant independent of L and M .

The proof of Proposition 2 is based on Lemma 1 and is provided in Appendix II-B.

Proposition 2 can be used to show that when $L < \frac{\log M}{I_{\max}}$, Problem (P) will have a trivial solution. The precise statement is given by the following corollary.

Corollary 1. Let $L < \frac{\log M}{I_{\max}}$, and suppose the decision maker has a uniform prior belief about the hypotheses. As $M \rightarrow \infty$, the optimal policy randomly guesses the true hypothesis without collecting any observation and hence, P_e , the probability of making a wrong declaration, approaches 1.

The proof of Corollary 1 is provided in Appendix II-C.

Corollary 2. For $L > \max\{1, \frac{\log M}{I_{\max}}\}$ and $\delta = \frac{1}{\log 2ML}$,

$$\underline{V}_2(\boldsymbol{\rho}) = \left[\frac{H(\boldsymbol{\rho})}{I_{\max}} + \frac{\log \frac{1-L^{-1}}{L^{-1}}}{D_{\max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1 - \frac{1}{\log 2ML}\}} - O(\log \log ML) \right]^+.$$

Remark 2. The lower bounds in Propositions 1 and 2 can be explained by the following intuition: For any measure of uncertainty $U(\cdot) : \mathbb{P}(\Theta) \rightarrow \mathbb{R}_+$, the number of samples required to reduce the uncertainty down to a target level $U(\boldsymbol{\rho}_{\text{target}})$ has to be at least $\frac{U(\boldsymbol{\rho}(0)) - U(\boldsymbol{\rho}_{\text{target}})}{\Delta_{\max}(U)}$, where $\Delta_{\max}(U)$ is the maximum amount of reduction in U associated with a single sample, i.e., $\Delta_{\max}(U) = \max_{a \in \mathcal{A}} \max_{\boldsymbol{\rho} \in \mathbb{P}(\Theta)} \{U(\boldsymbol{\rho}) - (\mathbb{T}^a U)(\boldsymbol{\rho})\}$. The lower bound in Proposition 1 is associated with such a lower bound when taking U to be the log-likelihood function, while the lower bound in Proposition 2 is associated with setting U to be the Shannon entropy.

Fact 2 (Proposition 3 in [21]). Under Assumptions 1 and 2 and for $L > 1$,

$$V^*(\boldsymbol{\rho}) \geq \left[\sum_{i=1}^M \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}} - \max_{j \neq i} \log \frac{\rho_i}{\rho_j}}{\max_{\lambda \in \Lambda(\mathcal{A})} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a)} - \frac{o(M \log L)}{\left(\max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \in \Omega} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a) \right)^2} \right]^+,$$

where $\frac{o(M \log L)}{M \log L} \rightarrow 0$ as $ML \rightarrow \infty$.

B. Upper Bounds for V^*

Next we propose two Markov policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$. Policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ have two operational phases. Phase 1 is the phase in which the belief about all hypotheses is below a certain threshold; while in phase 2, the belief about one of the hypotheses has passed that threshold and actions are selected in favor of that particular hypothesis. The difference between the two policies is in the actions they take in each phase.

Let μ_0 and η_0 be vectors in $\Lambda(\mathcal{A})$ such that

$$\begin{aligned}\mu_0 &:= \arg \max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \in \Omega} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a), \\ \eta_0 &:= \arg \max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \in \Omega} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a).\end{aligned}$$

For $i \in \Omega$, let μ_i and η_i be vectors in $\Lambda(\mathcal{A})$ such that

$$\begin{aligned}\mu_i &:= \arg \max_{\lambda \in \Lambda(\mathcal{A})} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a), \\ \eta_i &:= \arg \max_{\lambda \in \Lambda(\mathcal{A})} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a).\end{aligned}$$

Consider a threshold $\tilde{\rho}$, $\tilde{\rho} \in (\frac{1}{2}, 1 - L^{-1})$. Markov (randomized) policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are defined as follows.³

- If $\rho_i \geq 1 - L^{-1}$, retire and select H_i as the true hypothesis;
- If $\rho_i \in [\tilde{\rho}, 1 - L^{-1})$, then
 - $\tilde{\pi}_1(a|\rho) = \mu_{ia} \forall a \in \mathcal{A}$;
 - $\tilde{\pi}_2(a|\rho) = \eta_{ia} \forall a \in \mathcal{A}$;
- If $\rho_i \in [0, \tilde{\rho})$ for all $i \in \Omega$, then
 - $\tilde{\pi}_1(a|\rho) = \mu_{0a} \forall a \in \mathcal{A}$;
 - $\tilde{\pi}_2(a|\rho) = \eta_{0a} \forall a \in \mathcal{A}$.

Remark 3. Policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are equivalent for $M = 2$.

Remark 4. The performance of policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ depend highly on the problem at hand. Sections V-A and V-B will elaborate on this.

³Policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are not unique; they each represent a class of parameterized policies. The tilde in $\tilde{\pi}_1$ and $\tilde{\pi}_2$ is to emphasize the dependency of these policies on the threshold/parameter $\tilde{\rho}$.

Propositions 3 and 4 at the the end of this section provide two upper bounds \bar{V}_1 and \bar{V}_2 for the value function V^* . For notational simplicity, let

$$\begin{aligned} I_1 &:= \max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \in \Omega} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a), \\ I_2 &:= \max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \in \Omega} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a), \\ D_{\mu_i} &:= \max_{\lambda \in \Lambda(\mathcal{A})} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a), \\ D_{\eta_i} &:= \max_{\lambda \in \Lambda(\mathcal{A})} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a). \end{aligned}$$

It can be easily shown that $I_1, I_2, D_{\mu_i}, D_{\eta_i}, \forall i \in \Omega$, are non-increasing in M and $I_1 \leq D_{\mu_i} \leq D_{max}$ and $I_2 \leq D_{\eta_i} \leq D_{max}$ for all M . Note that Tables I and II provide respectively a list of the notations introduced in this section and their limiting values.

Proposition 3. *Under Assumptions 1 and 2, and for $L > 1$ and any $\rho \in \mathbb{P}_L(\Theta)$,*

$$\bar{V}_1(\rho) := \frac{H(\rho) + \log(M-1)}{I_1} + \sum_{i=1}^M \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k}}{D_{\mu_i}} + \frac{o(M + \log L)}{I_1^2},$$

is an upper bound for the optimal value function V^ where $\frac{o(M + \log L)}{M + \log L} \rightarrow 0$ as $ML \rightarrow \infty$.*

The proof is done by analyzing the performance of $\tilde{\pi}_1$ and is provided in Appendix III-A.

Proposition 4. *Under Assumptions 1 and 2, and for $L > 1$ and any $\rho \in \mathbb{P}_L(\Theta)$,*

$$\bar{V}_2(\rho) := \frac{H(\rho)}{I_2} + \sum_{i=1}^M \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}} + \frac{K_2''}{I_2}}{D_{\eta_i}} + 1,$$

is an upper bound for the optimal value function V^ and K_2'' is a constant independent of L and M .*

The proof is done by analyzing the performance of $\tilde{\pi}_2$ and is provided in Appendix III-B.

V. APPLICATIONS AND TECHNICAL CONSEQUENCES OF THE BOUNDS

In this section, we state and discuss the consequence of the bounds obtained in Section IV.

A. Order and Asymptotic Optimality

The lower and upper bounds provided in Section IV can be applied to establish the order optimality and asymptotic optimality of the proposed policies as defined below.

TABLE I
SUMMARY OF NOTATIONS

Notation	Description
I_{max}	$\max_{a \in \mathcal{A}} \max_{\hat{\rho} \in \mathbb{P}(\Theta)} I(\hat{\rho}; q_{\hat{\rho}}^a)$
D_{max}	$\max_{i,j \in \Omega} \max_{a \in \mathcal{A}} D(q_i^a q_j^a)$
I_1	$\max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \in \Omega} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a q_j^a)$
I_2	$\max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \in \Omega} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a \sum_{j \neq i} \frac{\hat{\rho}_j}{1-\hat{\rho}_i} q_j^a)$
D_{μ_i}	$\max_{\lambda \in \Lambda(\mathcal{A})} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a q_j^a)$
D_{η_i}	$\max_{\lambda \in \Lambda(\mathcal{A})} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a \sum_{j \neq i} \frac{\hat{\rho}_j}{1-\hat{\rho}_i} q_j^a)$

Definition 3. Let $V^\pi(\rho)$ denote the value function for policy π , i.e., the expected total cost achieved by policy π when the initial belief is ρ . For fixed M , policy π is referred to as *order optimal* in L if for all $\rho \in \mathbb{P}(\Theta)$,

$$\lim_{L \rightarrow \infty} \frac{V^\pi(\rho) - V^*(\rho)}{V^\pi(\rho)} < 1.$$

Definition 4. For fixed M , policy π is referred to as *asymptotically optimal of order-1* in L if for all $\rho \in \mathbb{P}(\Theta)$,

$$\lim_{L \rightarrow \infty} \frac{V^\pi(\rho) - V^*(\rho)}{V^\pi(\rho)} = 0.$$

Definition 5. For fixed M , policy π is referred to as *asymptotically optimal of order-2* in L if for all $\rho \in \mathbb{P}(\Theta)$, there exists a constant B independent of L such that

$$V^\pi(\rho) - V^*(\rho) \leq B.$$

Remark 5. It is clear from the definitions above that order optimality is weaker than asymptotic optimality of order-1; while asymptotic optimality of order-2 is the strongest notion. The notion of asymptotic optimality of order-1 was first introduced in [5] which naturally motivates the extension of higher orders.

Next definition extends the notions of order and asymptotic optimality defined above to the case where M increases as well.

TABLE II
SUMMARY OF LIMITING VALUES

Notation	Description	Notation	Description
\underline{I}_{max}	$\inf_M I_{max}$	\overline{I}_{max}	$\sup_M I_{max}$
\underline{D}_{max}	$\inf_M D_{max}$	\overline{D}_{max}	$\sup_M D_{max}$
\underline{I}_1	$\inf_M I_1$	\overline{I}_1	$\sup_M I_1$
\underline{I}_2	$\inf_M I_2$	\overline{I}_2	$\sup_M I_2$

Definition 6. Let ρ_u denote a uniform prior on the set of hypotheses. Policy π is referred to as *order optimal* and *asymptotically optimal of order-1* in L and M if respectively,

$$\lim_{L, M \rightarrow \infty} \frac{V^\pi(\rho_u) - V^*(\rho_u)}{V^\pi(\rho_u)} < 1, \quad \lim_{L, M \rightarrow \infty} \frac{V^\pi(\rho_u) - V^*(\rho_u)}{V^\pi(\rho_u)} = 0.$$

Next theorems establish order and asymptotic optimality of our proposed policies.

Theorem 1. Policy $\tilde{\pi}_1$ is asymptotically optimal of order-1 in L .

Proof: The proof simply follows from Definition 4, Fact 2, and Proposition 3. ■

Theorem 2. For $L > \frac{\log M}{I_{max}}$ and if $\underline{I}_2 > 0$, policy $\tilde{\pi}_2$ is order optimal in L and M .

Proof: The proof simply follows from Definition 6, Corollary 2, and Proposition 4. ■

Remark 6. Note that a sufficient condition for $\underline{I}_2 > 0$ can be obtained by strengthening Assumption 1 in the following manner: there exists $\zeta > 0$ such that for any hypothesis H_i , there exists an action $a \in \mathcal{A}$ for which $\min_{q \in Q_{-i}^a} D(q_i^a || q) \geq \zeta$ where for all $i \in \Omega, a \in \mathcal{A}, Q_{-i}^a$ is the convex hull of distributions $\{q_j^a(\cdot)\}_{j \in \Omega - \{i\}}$ on \mathcal{Z} .

Theorem 3. Policy $\tilde{\pi}_2$ attains asymptotic optimality of order-2 in L if

$$\min_{j \neq i} \max_{a \in \mathcal{A}} D(q_i^a || q_j^a) = D_{\eta_i}, \quad \forall i \in \Omega. \quad (5)$$

Furthermore, for $L > \frac{\log M}{I_{max}}$ and if $\overline{I}_{max} = \underline{I}_2$ and $D_{max} = D_{\eta_i}, \forall i \in \Omega$, policy $\tilde{\pi}_2$ is asymptotically optimal of order-1 in L and M .

Proof: The proof of the first part follows from Definition 5, Proposition 1, and Proposition 4. The proof of the second part follows from Definition 6, Proposition 4, and Corollary 2. ■

B. Information Acquisition Rate and Reliability

In this section, we explain the primal (constrained) version of Problem (P), referred to as Problem (P'), and use the obtained bounds to extend the notions of achievable (communication) rate and error exponent to the context of active sequential hypothesis testing. The proofs of all the results in this section are provided in Appendix V.

Problem (P') [Information Acquisition Problem]

Consider a hypothesis testing problem with M hypotheses of interest, action space \mathcal{A} , and observation kernels $\{q_i^a(\cdot)\}_{i \in \Omega, a \in \mathcal{A}}$. A Bayesian decision maker with prior belief $\boldsymbol{\rho}(0)$ is responsible to find the true hypothesis with the objective to

$$\text{minimize } \mathbb{E}[\tau] \quad \text{subject to } \text{Pe} \leq \epsilon, \quad (6)$$

where τ is the stopping time at which the decision maker retires, Pe is the probability of making a wrong declaration, and $\epsilon > 0$ denotes the desired probability of error.

Problem (P) can be viewed as a Lagrangian relaxation of Problem (P'). It is somewhat intuitive that as $L \rightarrow \infty$ the solution of Problem (P) is closely related to that of Problem (P') when $\epsilon \rightarrow 0$. The following lemma makes this intuition precise.

Lemma 2. *Let $\mathbb{E}[\tau_\epsilon^*]$ denote the minimum expected number of samples required to achieve $\text{Pe} \leq \epsilon$. We have*

$$\mathbb{E}[\tau_\epsilon^*] \geq (1 - \epsilon L) (V^*(\boldsymbol{\rho}(0)) - 1), \quad (7)$$

where $V^*(\boldsymbol{\rho}(0))$ is the optimal solution to Problem (P) for prior belief $\boldsymbol{\rho}(0)$ and penalty of wrong declaration L .

Let $\mathbb{E}^\pi[\tau]$ and Pe^π denote respectively the expected stopping time (or equivalently the expected number of collected samples) and the probability of error under policy π . Following the notations in [38], we define $M^\pi(t, \epsilon)$ as the maximum number of hypotheses among which policy π can find the true hypothesis with $\mathbb{E}^\pi[\tau] \leq t$ and $\text{Pe}^\pi \leq \epsilon$. Policy π is said to achieve information acquisition rate $R > 0$ with reliability (also known as error exponent) $E > 0$ if

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log M^\pi(t, 2^{-Et}) = R. \quad (8)$$

For fixed number of hypotheses, hence at information rate $R = 0$, policy π is said to achieve reliability $E > 0$ if

$$\lim_{t \rightarrow \infty} \frac{-1}{t} \log \text{Pe}^\pi(t, M) = E, \quad (9)$$

where $\text{Pe}^\pi(t, M)$ is the smallest probability of error that policy π can guarantee when looking for the true hypothesis among M hypotheses with $\mathbb{E}^\pi[\tau] \leq t$.

The reliability function $E(R)$ is defined as the maximum achievable error exponent at information rate R . Next we use the bounds obtained in Section IV as well as Lemma 2 to characterize upper and lower bounds on the maximum achievable information acquisition rate and the optimal reliability function.

Before we proceed, we refer the reader to Tables I, II for the list of notations introduced in Section IV. Also let D_1 and D_2 denote respectively the harmonic mean of $\{D_{\mu_i}\}_{i \in \Omega}$ and $\{D_{\eta_i}\}_{i \in \Omega}$, i.e.,

$$D_1 = M \left(\sum_{i=1}^M \frac{1}{D_{\mu_i}} \right)^{-1}, \quad D_2 = M \left(\sum_{i=1}^M \frac{1}{D_{\eta_i}} \right)^{-1}.$$

Moreover, let

$$\underline{D}_1 = \inf_M D_1, \quad \underline{D}_2 = \inf_M D_2.$$

Corollary 3. *Suppose the hypotheses are equiprobable, i.e., $P(\{\theta = i\}) = \frac{1}{M}$, $\forall i \in \Omega$. No policy can achieve positive reliability $E > 0$ at rates higher than \bar{I}_{max} . Furthermore,*

$$E(R) \leq \begin{cases} \bar{D}_{max} \left(1 - \frac{R}{\bar{I}_{max}} \right) & R \in (0, \bar{I}_{max}) \\ D_1 & R = 0 \end{cases} \quad (10)$$

Remark 7. Corollary 3 establishes an upper bound, \bar{I}_{max} , on maximum achievable information acquisition rate. As shown in Appendix V-C, this can be strengthened to show that no policy can achieve diminishing error probability at rates higher than \bar{I}_{max} .

Corollary 4. *For fixed M , hence at information rate $R = 0$, a policy π can achieve the maximum reliability, i.e. $E = D_1$, if and only if it is asymptotically optimal in L .*

Corollary 4 implies that for fixed M , hence at $R = 0$, policies $\tilde{\pi}_1$ and π^* achieve the optimal error exponent.

Corollary 5. *A policy π can achieve a non-zero rate $R > 0$ with non-zero reliability $E > 0$ only if it is order optimal in L and M .*

Corollary 6. *Suppose the hypotheses are equiprobable, i.e., $P(\{\theta = i\}) = \frac{1}{M}$, $\forall i \in \Omega$. Policy $\tilde{\pi}_2$ can achieve any rate $R \in [0, \underline{I}_2]$ with reliability E if*

$$E \leq \underline{D}_2 \left(1 - \frac{R}{\underline{I}_2} \right). \quad (11)$$

Fig. 1 summarizes the results above. The upper bound on the reliability function is shown in red. Policy $\tilde{\pi}_1$ achieves the optimal reliability D_1 at $R = 0$ with no provable guarantee for $R > 0$ (this point is shown in green); while policy $\tilde{\pi}_2$ ensures an exponentially decaying error probability (the error exponent is shown in blue) for $R \in [0, \underline{I}_2]$.

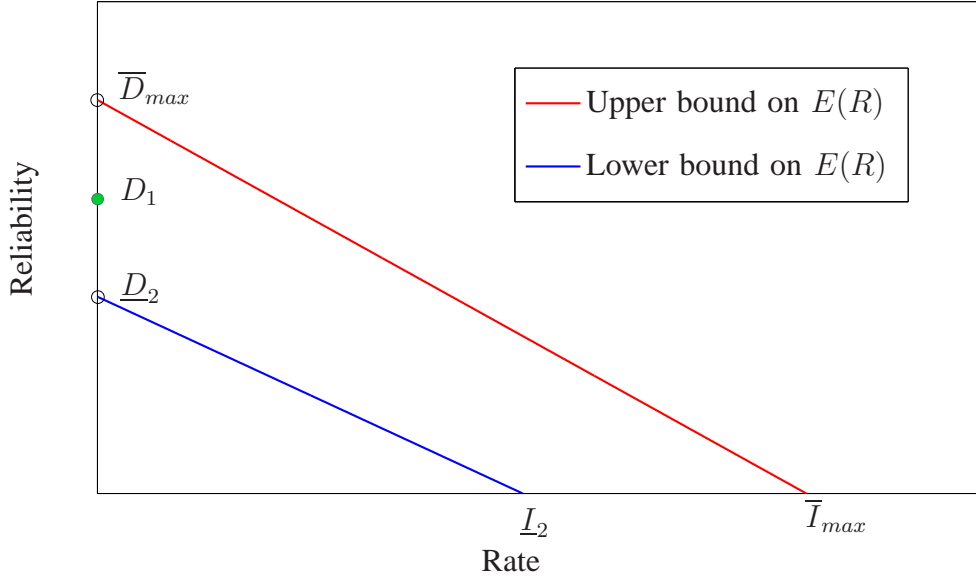


Fig. 1. Lower and upper bounds on the optimal reliability function $E(R)$.

Remark 8. It can be shown that any optimal policy π^* for Problem (P) also achieves any rate $R \in [0, \underline{I}_2]$ with reliability E satisfying (11).

C. Adaptivity Gain

In this section, we first define a class of policies which do not fully utilize the observation outcomes. We then discuss the performance gap between these policies and the optimal one.

Definition 7. A policy π is referred to as *open-loop* or *non-adaptive* if under which, sensing actions are selected independent of the observation outcomes (hence independent of the belief state). For a given vector $\lambda \in \Lambda(\mathcal{A})$, non-adaptive policy π_λ selects sensing actions $a \in \mathcal{A}$ with probability $\pi_\lambda(a|\rho) = \lambda_a$ independent of ρ until the stopping time τ is reached.

Definition 8. Let $V_{\lambda^*}(\rho)$ denote the value function for the best (randomized) non-adaptive policy. *Adaptivity gain* is defined as $V_{\lambda^*}(\rho) - V^*(\rho)$, i.e., the increase in the expected total cost under the best non-adaptive policy relative to the optimal policy.

The advantage of non-adaptive policies is that they do not require knowledge of the observations and/or a careful reevaluation of the belief state when selecting sensing actions. The adaptivity gain defined above characterizes the loss in the performance due to non-adaptive selection of sensing actions. Next we show that for fixed M , the performance gap between the non-adaptive policy and $\tilde{\pi}_1$ (hence the optimal one) grows at least logarithmically as the penalty L increases.

Let $V_\lambda(\rho)$ denote the expected total cost when the initial belief state is ρ and non-adaptive policy π_λ is enforced. The following proposition provides a lower bound for $V_\lambda(\rho)$.

Proposition 5. *Under Assumption 1 and for $L > 1$, $V_\lambda(\rho)$ is lower bounded by*

$$\underline{V}_\lambda(\rho) = \left[\sum_{i=1}^M \rho_i \max_{j \neq i} \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i}{\rho_j}}{\sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a)} - K'_\lambda \right]^+,$$

where K'_λ is independent of penalty L .

The proof is very similar to the proof of Proposition 1 and is provided in Appendix IV.

Corollary 7. *Unless there exists a $\hat{\lambda} \in \Lambda(\mathcal{A})$ such that,*

$$\min_{j \neq i} \sum_{a \in \mathcal{A}} \hat{\lambda}_a D(q_i^a || q_j^a) = \max_{\lambda \in \Lambda(\mathcal{A})} \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a), \quad \forall i \in \Omega, \quad (12)$$

the adaptivity gain grows at least logarithmically with L .

Corollary 7 can be further simplified for the binary case ($M = 2$) as follows.

Corollary 8. *In the active binary hypothesis testing, the adaptivity gain grows logarithmically in L if*

$$\arg \max_{a \in \mathcal{A}} D(q_1^a || q_2^a) \neq \arg \max_{a \in \mathcal{A}} D(q_2^a || q_1^a).$$

VI. EXAMPLES

In this section, we consider important special cases of the active hypothesis testing under which conditions of Theorem 3 hold, hence establishing the asymptotic optimality of $\tilde{\pi}_2$.

A. Binary Hypothesis Testing ($M = 2$)

Consider Problem (P) for $M = 2$. In this setting, as noted in Remark 3, policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are equivalent and by Theorem 1, both policies are asymptotically optimal of order-1 in L . Asymptotic optimality of order-2 of $\tilde{\pi}_1$ and $\tilde{\pi}_2$ is also verified from Theorem 3 since equality (5) holds trivially for $M = 2$. Furthermore, we obtain

$$V^*(\rho) = \rho_1 \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_1}{\rho_2}}{\max_{a \in \mathcal{A}} D(q_1^a || q_2^a)} + \rho_2 \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_2}{\rho_1}}{\max_{a \in \mathcal{A}} D(q_2^a || q_1^a)} \pm O(1).$$

The problem of reliability (error exponent) associated with passive binary hypothesis testing with fixed-length (non-sequential) as well as variable-length (sequential) sample size has been studied by [39]–[41]. Recently, the authors in [42], [43] have generalized this problem for fixed-length and variable-length sample size respectively, to the active testing context. Our work complements the findings in [43] by providing an asymptotic optimal solution in a total cost (and Bayesian) sense.

B. Noisy Dynamic Search

Consider the problem of sequentially searching for one and only object of interest in M locations where the goal is to find the object quickly and accurately. Let $\mathcal{A} = 2^\Omega$ be the set of all allowable combinations of locations that can be searched in one time slot. The outcome of the search is a random variable with probability density function f_{obj} if the object is in the location(s) that are being searched; otherwise, it is distributed as f_{noise} .

The problem above can be modeled as an active hypothesis testing with action space \mathcal{A} and the observation kernels

$$q_i^a(\cdot) = \begin{cases} f_{\text{obj}}(\cdot) & \text{if } i \in a \\ f_{\text{noise}}(\cdot) & \text{if } i \notin a \end{cases}, \quad \forall i \in \Omega, \forall a \in \mathcal{A}.$$

Lemma 3. *Consider the problem of noisy dynamic search explained above. Under the symmetry condition $f_{\text{obj}}(z) = f_{\text{noise}}(b - z)$ for some $b \in \mathbb{R}$,*

$$\underline{I}_2 = D(f_{\text{obj}} || \frac{1}{2}f_{\text{obj}} + \frac{1}{2}f_{\text{noise}}) = \bar{I}_{\max}, \quad (13)$$

$$\min_{j \neq i} \max_{a \in \mathcal{A}} D(q_i^a || q_j^a) = D_{\eta_i} = D(f_{\text{obj}} || f_{\text{noise}}) = D_{\max}, \quad \forall i \in \Omega. \quad (14)$$

Lemma 3 together with Theorem 3 implies that policy $\tilde{\pi}_2$ attains 1) asymptotic optimality of order-2 in L ; and 2) asymptotic optimality of order-1 in L and M .

Under the condition of Lemma 3, the schemes proposed in [5], [22], [23] simplify to the one that searches, at each instant, a location with the highest probability of having the object. This scheme, which was also studied in [44] in a finite horizon context with symmetric observations and was shown to be optimal among policies that can only search a single location at a time, has an information acquisition rate that is restricted to zero; while at zero rate, it achieves asymptotic optimality and maximum error exponent \bar{D}_{\max} . In contrast, in [45], [46], a generalization of binary search was proposed in which the locations are partitioned along the median of the posterior and, in effect, are searched along a generalized binary tree. It was shown in [45], [46] that in the special case of Bernoulli noise where $f_{\text{obj}} = \mathcal{B}(1 - p)$, and $f_{\text{noise}} = \mathcal{B}(p)$, $p \in (0, 0.5)$, the proposed policy can achieve any rate

$R < D(f_{\text{obj}} || \frac{1}{2}f_{\text{obj}} + \frac{1}{2}f_{\text{noise}})$ with reliability $E = D(f_{\text{obj}} || \frac{1}{2}f_{\text{obj}} + \frac{1}{2}f_{\text{noise}}) - R$. In other words, the policy was shown to be asymptotic optimal in M (since $\underline{I}_2 = \bar{I}_{\max}$) but only order optimal in L (since $0 < D(f_{\text{obj}} || \frac{1}{2}f_{\text{obj}} + \frac{1}{2}f_{\text{noise}}) < D(f_{\text{obj}} || f_{\text{noise}}) = \bar{D}_{\max}$).

Our proposed policy $\tilde{\pi}_2$ combines the best of the above two worlds: in its first phase, by randomly selecting actions from \mathcal{A} , it ensures maximum acquisition rate $D(f_{\text{obj}} || \frac{1}{2}f_{\text{obj}} + \frac{1}{2}f_{\text{noise}})$ obtained by the generalized binary search of [45], [46]; while its second phase coincides with the schemes in [5], [22], [23] ensuring the maximum feasible error exponent.

C. Variable-Length Coding with Noiseless Feedback

Consider the feedback communication system depicted in Fig. 2. Let $\Omega = \{1, 2, \dots, M\}$ denote the message set. We assume a sender wishes to communicate a discrete message $\theta \in \Omega$ to the receiver with probability of error less than some $\epsilon > 0$ over a noisy memoryless channel. The channel is described by finite input set \mathcal{X} and output set \mathcal{Y} (possibly uncountable), and a collection of conditional probabilities $P(Y|X)$. Let C denote the Shannon capacity of this channel and C_1 the KL divergence between its two most distinguishable inputs:

$$C_1 = \max_{x, x' \in \mathcal{X}} D(P(Y|X=x) || P(Y|X=x')).$$

The encoder receives a perfect knowledge about the decoder's past received signals through a noiseless causal feedback link. Using this knowledge, encoder decides what to transmit and when to terminate the transmission. The objective is to find encoding/stopping/decoding rules which achieve a desired probability of error, i.e., ϵ , with minimum expected number of channel uses. To achieve this goal, the encoder sequentially and causally selects input sequence $\{X_t\}$ until a stopping time τ , and the decoder follows a decoding rule $d : \mathcal{Y}^\tau \rightarrow \Omega$. The objective is to ensure that $P(\{\hat{\theta} \neq \theta\}) \leq \epsilon$ while $\mathbb{E}[\tau]$ is minimized, where $\hat{\theta} = d(Y^\tau)$.

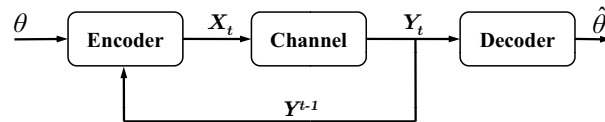


Fig. 2. A noisy memoryless channel with a noiseless causal feedback link.

Let $\mathcal{E} := \{e(\cdot) : \Omega \rightarrow \mathcal{X}\}$ be the set of all mappings from the set of messages Ω to \mathcal{X} . In [47], using the results from [48], we showed that without loss of generality a fictitious agent can be added to this communication system who has access to the past channel outputs and is responsible for selecting

actions from $\mathcal{E} \cup \{d\}$. The choice of decoding, i.e., action d , marks the termination of the transmission phase, the stopping time τ , while the choice of encoding function e_t at time t determines the input to the channel at time t , i.e., $X_t = e_t(\theta)$. The addition of this fictitious agent does not change the nature of the problem. The reason is that the decision of the fictitious agent at any time t solely relies on Y^{t-1} which is fully observable by both transmitter and receiver and hence are easily replicated at transmitter and receiver in isolate.

From this point of view, the problem of variable-length coding with noiseless feedback is closely related to a special case of Problem (P') defined in Section V-B where a fictitious agent plays the role of the Bayesian decision maker whose K available actions coincide with \mathcal{E} ($\mathcal{A} = \mathcal{E}$ and $K = |\mathcal{X}|^M$), and whose observation kernels are given by $q_i^e(y) = P(Y = y|X = e(i))$.

Lemma 4. *For the problem of variable-length coding with feedback we have,*

$$\underline{I}_2 = C = \overline{I}_{max}, \quad (15)$$

$$\underline{D}_2 = C_1 = \overline{D}_{max}. \quad (16)$$

The proof is provided in Appendix VI.

Therefore, for the problem of variable-length coding with feedback with equiprobable messages, i.e., $P(\{\theta = i\}) = \frac{1}{M}$, $\forall i \in \Omega$, we have

$$\frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} - O(\log \log \frac{M}{\epsilon}) \leq \mathbb{E}[\tau_\epsilon^*] \leq \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} + O(1). \quad (17)$$

Furthermore, the optimal reliability function can be characterized as

$$E(R) = C_1 \left(1 - \frac{R}{C}\right), \quad (18)$$

and is achieved by policy $\tilde{\pi}_2$.

The problem above was first tackled by Burnashev in his seminal paper [11]. In fact, policy $\tilde{\pi}_2$ is nothing but a natural generalization of Burnashev's two-phase coding scheme.

The lower bound in (17) was proved in [11] using a lengthy Martingale argument, and later was reproved in [49] and [38]. The proofs in [49] and [38] parallel the two-phase encoding scheme corresponding to the upper bound but provides slightly tighter lower bound (the double logarithmic term can be replaced by a constant when M is fixed). In this light, Proposition 2 provides yet another alternative proof for Burnashev's lower bound utilizing the MDP formulation (4).

VII. DISCUSSION AND FUTURE WORK

In this paper, we considered the problem of active sequential M -ary hypothesis testing. Using a DP formulation, we characterized the optimal value function V^* . Three lower bounds (complementary for

various values of the parameters of the problem) were obtained for the optimal value function V^* . We also proposed two heuristic policies whose performance analysis resulted in two upper bounds for V^* . Subsequently, we discussed important consequences of the bounds and established order and asymptotic optimality of the proposed policies under different scenarios. An important problem which remains is further improvement of the performance bounds.

In this paper, we focused on sequential policies, i.e., policies whose sample size is not known initially and is dependent on the observation outcomes. There exist other types of policies in the literature. For example, non-sequential policies take a fixed number of samples (independent of observation outcomes) and make the final decision afterwards; while multi-stage policies (introduced in [33], [34]) can take a retire/declare action only at the end of each stage, and stages are not necessarily of the same size. Comparing the performance of sequential, non-sequential, and multi-stage policies in the context of active hypothesis testing is an area of future work.

APPENDIX I

PROOF OF LEMMA 1

To prove Lemma 1, we have to slightly modify the state space and introduce new notations. We assume that after taking the retire/declare action, the system goes to the termination state, denoted by F , and remains in that state for the rest of the time. The state space is modified to $\mathcal{S} = \mathbb{P}(\Theta) \cup \{F\}$ to include the termination state. For $a \in \mathcal{A} \cup \{d\}$, $s \in \mathcal{S}$, let

$$c^a(s) = \begin{cases} 1 & \text{if } s = \boldsymbol{\rho} \in \mathbb{P}(\Theta), a \in \mathcal{A} \\ \min_{j \in \Omega} (1 - \rho_j) L & \text{if } s = \boldsymbol{\rho} \in \mathbb{P}(\Theta), a = d \\ 0 & \text{if } s = F \end{cases}.$$

The Bayes operator is modified as follows:

$$\Phi^a(s, z) = \begin{cases} \Phi^a(\boldsymbol{\rho}, z) & \text{if } s = \boldsymbol{\rho} \in \mathbb{P}(\Theta), a \in \mathcal{A} \\ F & \text{if } s = \boldsymbol{\rho} \in \mathbb{P}(\Theta), a = d \\ F & \text{if } s = F \end{cases}.$$

Using the notations above, the condition $V(\boldsymbol{\rho}) \leq \min\{1 + \min_{a \in \mathcal{A}} (\mathbb{T}^a V)(\boldsymbol{\rho}), \min_{j \in \Omega} (1 - \rho_j) L\}$ is rewritten as

$$\begin{aligned} V(F) &= 0, \\ V(s) &\leq \min_{a \in \mathcal{A} \cup \{d\}} \{c^a(s) + \mathbb{E}[V(\Phi^a(s, Z))]\}, \forall s \in \mathcal{S} - \{F\}. \end{aligned} \tag{19}$$

Let S_0, S_1, S_2, \dots be a sequence of random variables denoting the belief states at times $t = 0, 1, 2, \dots$ starting from belief state s , i.e.,

$$\begin{aligned} S_0 &= s, \\ S_n &= \Phi^{A(n-1)}(S_{n-1}, Z), \quad \forall n, n > 0. \end{aligned}$$

Using (19) iteratively for N times, we obtain

$$\begin{aligned} V(s) &\leq \mathbb{E}^{\pi^*}[c^{A(0)}(s)] + \mathbb{E}^{\pi^*}[V(\Phi^{A(0)}(s, Z))] \\ &= \mathbb{E}^{\pi^*}[c^{A(0)}(S_0)] + \mathbb{E}^{\pi^*}[V(S_1)] \\ &\leq \mathbb{E}^{\pi^*}\left[\sum_{n=0}^1 c^{A(n)}(S_n)\right] + \mathbb{E}^{\pi^*}[V(S_2)] \\ &\leq \mathbb{E}^{\pi^*}\left[\sum_{n=0}^{N-1} c^{A(n)}(S_n)\right] + \mathbb{E}^{\pi^*}[V(S_N)], \end{aligned}$$

where superscript π^* implies that actions are selected according to an optimal policy π^* . Taking the limit as $N \rightarrow \infty$, we obtain

$$\begin{aligned} V(s) &\stackrel{(a)}{\leq} \mathbb{E}^{\pi^*}\left[\sum_{n=0}^{\infty} c^{A(n)}(S_n)\right] + \liminf_{N \rightarrow \infty} \mathbb{E}^{\pi^*}[V(S_N)] \\ &\stackrel{(b)}{=} V^*(s) + \liminf_{N \rightarrow \infty} \mathbb{E}^{\pi^*}[V(S_N)] \\ &= V^*(s) + \liminf_{N \rightarrow \infty} \mathbb{E}^{\pi^*}[V(F)\mathbf{1}_{\{S_N=F\}} + V(S_N)\mathbf{1}_{\{S_N \neq F\}}] \\ &= V^*(s) + \liminf_{N \rightarrow \infty} \mathbb{E}^{\pi^*}[V(S_N)\mathbf{1}_{\{S_N \neq F\}}] \\ &\stackrel{(c)}{\leq} V^*(s) + L \liminf_{N \rightarrow \infty} P^{\pi^*}(S_N \neq F) \\ &\stackrel{(d)}{=} V^*(s), \end{aligned}$$

where (a) follows from the monotone convergence theorem, (b) follows from the definition of V^* , (c) follows from the fact that for any $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$, $V(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j)L \leq L$, and (d) holds since $L \geq V^*(s) \geq \mathbb{E}^{\pi^*}[\tau] = \sum_{n=0}^{\infty} P^{\pi^*}(\tau > n) = \sum_{n=0}^{\infty} P^{\pi^*}(S_n \neq F)$.

APPENDIX II

PROOF OF PROPOSITIONS 1 AND 2 AND COROLLARY 1

A. Proof of Proposition 1

Let Γ be the set of all mappings $\gamma : \Omega \rightarrow \Omega$ such that $\gamma(i) \neq i$ for $i \in \Omega$. Now associated with any $\gamma \in \Gamma$ define

$$\underline{V}_1^\gamma(\boldsymbol{\rho}) = \left[\sum_{i=1}^M \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i}{\rho_{\gamma(i)}}}{\max_{\hat{a} \in \mathcal{A}} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} - K'_1 \right]^+.$$

Next we use Lemma 1 to show that $V^* \geq \underline{V}_1^\gamma$ for all $\gamma \in \Gamma$. In particular, we show that for all $\gamma \in \Gamma$ and all $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$, $\underline{V}_1^\gamma(\boldsymbol{\rho}) \leq \min \{1 + \min_{a \in \mathcal{A}} (\mathbb{T}^a \underline{V}_1^\gamma)(\boldsymbol{\rho}), \min_{j \in \Omega} (1 - \rho_j)L\}$. For any $\boldsymbol{\rho}$ such that $\underline{V}_1^\gamma(\boldsymbol{\rho}) = 0$, the inequality holds trivially. For $\underline{V}_1^\gamma(\boldsymbol{\rho}) > 0$ and for any action $a \in \mathcal{A}$ we have

$$\begin{aligned} & (\mathbb{T}^a \underline{V}_1^\gamma)(\boldsymbol{\rho}) \\ & \geq \sum_{i=1}^M \int \rho_i q_i^a(z) \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i q_i^a(z)}{\rho_{\gamma(i)} q_{\gamma(i)}^a(z)}}{\max_{\hat{a} \in \mathcal{A}} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} dz - K'_1 \\ & = \underline{V}_1^\gamma(\boldsymbol{\rho}) - \sum_{i=1}^M \rho_i \frac{D(q_i^a \| q_{\gamma(i)}^a)}{\max_{\hat{a} \in \mathcal{A}} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} \\ & \geq \underline{V}_1^\gamma(\boldsymbol{\rho}) - 1. \end{aligned}$$

Claim 1 (in Appendix VII). Constant K'_1 can be selected independent of L such that $\underline{V}_1^\gamma(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j)L$ is satisfied for all $\gamma \in \Gamma$.

Using Claim 1 and letting $\underline{V}_1(\cdot) = \max_{\gamma \in \Gamma} \underline{V}_1^\gamma(\cdot)$, we have the assertion of the proposition.

B. Proof of Proposition 2

Recall that $I_{max} = \max_{a \in \mathcal{A}} \max_{\hat{\boldsymbol{\rho}} \in \mathbb{P}(\Theta)} I(\hat{\boldsymbol{\rho}}; q_{\hat{\boldsymbol{\rho}}}^a)$ and $\alpha(L, M) = \frac{M-1}{M-1+2LI_{max}}$. We first show that for all belief vectors $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$,

$$V^*(\boldsymbol{\rho}) \geq \left[\frac{H(\boldsymbol{\rho}) - H([\alpha(L, M), 1 - \alpha(L, M)]) - \alpha(L, M) \log(M-1)}{I_{max}} + \alpha(L, M)L \right]^+. \quad (20)$$

Note that the right-hand side of (20) can be written as

$$G(\boldsymbol{\rho}) := \left[\frac{H(\boldsymbol{\rho}) - H(\boldsymbol{\nu})}{I_{max}} + \alpha(L, M)L \right]^+,$$

where

$$\boldsymbol{\nu} = \left[\frac{\alpha(L, M)}{M-1}, \dots, \frac{\alpha(L, M)}{M-1}, 1 - \alpha(L, M) \right].$$

Next we show that for all $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$, $G(\boldsymbol{\rho}) \leq \min \left\{ 1 + \min_{a \in \mathcal{A}} (\mathbb{T}^a G)(\boldsymbol{\rho}), \min_{j \in \Omega} (1 - \rho_j)L \right\}$. For any $\boldsymbol{\rho}$ such that $G(\boldsymbol{\rho}) = 0$, the inequality holds trivially. For $G(\boldsymbol{\rho}) > 0$ and for any action $a \in \mathcal{A}$ we have

$$\begin{aligned} (\mathbb{T}^a G)(\boldsymbol{\rho}) &= \frac{\int H(\Phi^a(\boldsymbol{\rho}, z)) q_{\boldsymbol{\rho}}^a(z) dz - H(\boldsymbol{\nu})}{I_{max}} + \alpha(L, M)L \\ &= \frac{H(\boldsymbol{\rho}) - I(\boldsymbol{\rho}; q_{\boldsymbol{\rho}}^a) - H(\boldsymbol{\nu})}{I_{max}} + \alpha(L, M)L \\ &= G(\boldsymbol{\rho}) - \frac{I(\boldsymbol{\rho}; q_{\boldsymbol{\rho}}^a)}{I_{max}} \\ &\geq G(\boldsymbol{\rho}) - 1, \end{aligned}$$

where the last inequality follows from the fact that $I(\boldsymbol{\rho}; q_{\boldsymbol{\rho}}^a) \leq \max_{\hat{a} \in \mathcal{A}} \max_{\hat{\boldsymbol{\rho}} \in \mathbb{P}(\Theta)} I(\hat{\boldsymbol{\rho}}; q_{\hat{\boldsymbol{\rho}}}^{\hat{a}}) = I_{max}$. Therefore,

$$G(\boldsymbol{\rho}) \leq 1 + \min_{a \in \mathcal{A}} (\mathbb{T}^a G)(\boldsymbol{\rho}).$$

What remains is to show that $G(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j)L$. Rewriting G as

$$G(\boldsymbol{\rho}) = \left[\frac{\sum_{i=1}^{M-1} \rho_i \log \frac{1}{\rho_i} + (1 - \sum_{i=1}^{M-1} \rho_i) \log \frac{1}{1 - \sum_{i=1}^{M-1} \rho_i} - H(\boldsymbol{\nu})}{I_{max}} + \alpha(L, M)L \right]^+,$$

we can compute the gradient at $\boldsymbol{\nu}$. For all $i = 1, 2, \dots, M-1$,

$$\begin{aligned} \frac{\partial G}{\partial \rho_i}(\boldsymbol{\nu}) &= \left(\log \frac{1}{\rho_i} - \log e - \log \frac{1}{1 - \sum_{i=1}^{M-1} \rho_i} + \log e \right) / I_{max} \Big|_{\boldsymbol{\rho}=\boldsymbol{\nu}} \\ &= \left(\log \frac{\rho_M}{\rho_i} \right) / I_{max} \Big|_{\boldsymbol{\rho}=\boldsymbol{\nu}} = \left(\log \frac{1 - \alpha(L, M)}{\frac{\alpha(L, M)}{M-1}} \right) / I_{max} = L. \end{aligned}$$

Furthermore, $G(\boldsymbol{\nu}) = \alpha(L, M)L = (1 - \nu_M)L$. Without loss of generality and since both functions $G(\boldsymbol{\rho})$ and $\min_{j \in \Omega} (1 - \rho_j)L$ are symmetric, let us focus on $\mathbb{P}_M(\Theta) := \{\boldsymbol{\rho} \in \mathbb{P}(\Theta) : \rho_M \geq \rho_i, \forall i \in \Omega - \{M\}\}$. In this case, $\min_{j \in \Omega} (1 - \rho_j)L = (1 - \rho_M)L = \sum_{i=1}^{M-1} \rho_i L$ and hence, $\min_{j \in \Omega} (1 - \rho_j)L$ is the tangent hyperplane to $G(\boldsymbol{\rho})$ at $\boldsymbol{\nu}$. This along with concavity of function G implies $G(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j)L$. Using Lemma 1, we have the assertion of the proposition.

Next we need to show that

$$\begin{aligned} V^*(\boldsymbol{\rho}) \geq \underline{V}_2(\boldsymbol{\rho}) &= \left[\frac{H(\boldsymbol{\rho}) - H([\delta, 1 - \delta]) - \delta \log(M-1)}{I_{max}} \right. \\ &\quad \left. + \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - K'_2 \right]^+. \end{aligned} \quad (21)$$

We show this in two steps. First we consider the following concave function:

$$J'(\boldsymbol{\rho}) := \left[\sum_{i=1}^M \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}} + \log \xi - \log \frac{\rho_i}{1-\rho_i}}{D_{max}} - K'_2 \right]^+.$$

We use Jensen's inequality to show that

Claim 2 (in Appendix VII). For all $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$, $J'(\boldsymbol{\rho}) \leq 1 + \min_{a \in \mathcal{A}}(\mathbb{T}^a J')(\boldsymbol{\rho})$.

Next we define $J(\boldsymbol{\rho}) = \max\{J'(\boldsymbol{\rho}), J''(\boldsymbol{\rho})\}$ where $J''(\boldsymbol{\rho})$ is the right-hand side of (21), i.e.,

$$J''(\boldsymbol{\rho}) = \left[\frac{H(\boldsymbol{\rho}) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{max}} + \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - K'_2 \right]^+.$$

- **Case 1:** For all $\boldsymbol{\rho}$ such that $J(\boldsymbol{\rho}) = 0$ or $J(\boldsymbol{\rho}) = J'(\boldsymbol{\rho})$, it is trivial from Claim 2 that

$$J(\boldsymbol{\rho}) = J'(\boldsymbol{\rho}) \leq 1 + \min_{a \in \mathcal{A}}(\mathbb{T}^a J')(\boldsymbol{\rho}) \leq 1 + \min_{a \in \mathcal{A}}(\mathbb{T}^a J)(\boldsymbol{\rho}). \quad (22)$$

- **Case 2:** For all $\boldsymbol{\rho}$ such that $J(\boldsymbol{\rho}) = J''(\boldsymbol{\rho}) > 0$, and for any action $a \in \mathcal{A}$, we have

$$\begin{aligned} (\mathbb{T}^a J)(\boldsymbol{\rho}) &= \int J(\Phi^a(\boldsymbol{\rho}, z)) q_{\boldsymbol{\rho}}^a(z) dz \\ &\stackrel{(a)}{\geq} \frac{\int H(\Phi^a(\boldsymbol{\rho}, z)) q_{\boldsymbol{\rho}}^a(z) dz - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{max}} \\ &\quad + \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - K'_2 \\ &= J''(\boldsymbol{\rho}) - \frac{I(\boldsymbol{\rho}; q_{\boldsymbol{\rho}}^a)}{I_{max}} \\ &\geq J''(\boldsymbol{\rho}) - 1 \\ &\stackrel{(b)}{=} J(\boldsymbol{\rho}) - 1, \end{aligned} \quad (23)$$

where (a) follows from Claim 3 below and (b) holds since $\boldsymbol{\rho}$ is such that $J(\boldsymbol{\rho}) = J''(\boldsymbol{\rho})$.

Claim 3 (in Appendix VII). Let $\boldsymbol{\rho}$ be such that $J(\boldsymbol{\rho}) = J''(\boldsymbol{\rho}) > 0$. If Assumption 2 holds, then for all actions $a \in \mathcal{A}$ and observations $z \in \mathcal{Z}$,

$$J(\Phi^a(\boldsymbol{\rho}, z)) \geq \frac{H(\Phi^a(\boldsymbol{\rho}, z)) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{max}} + \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - K'_2. \quad (24)$$

In other words, from (22) and (23), we have that

$$J(\boldsymbol{\rho}) \leq 1 + \min_{a \in \mathcal{A}}(\mathbb{T}^a J)(\boldsymbol{\rho}). \quad (25)$$

We also have

Claim 4 (in Appendix VII). For $L > \frac{\log M}{I_{max}}$, constant K'_2 can be selected independent of L and M such that $J(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j) L$ is satisfied.

Lemma 1, together with (25) and Claim 4, implies that $V^* \geq J = \max\{J', J''\} \geq J'' = \underline{V}_2$. This is a slightly stronger result than (21).

C. Proof of Corollary 1

We first show that for all belief vectors $\boldsymbol{\nu} \in \mathbb{P}(\Theta)$ for which $\nu_i = 1 - \alpha(L, M)$ for an $i \in \Omega$ and $\nu_j = \frac{\alpha(L, M)}{M-1}$ for all $j \in \Omega - \{i\}$, the optimal action is to retire and declare H_i as the true hypothesis, i.e., $V^*(\boldsymbol{\nu}) = L(1 - \nu_i)$. Without loss of generality, consider $i = M$, hence

$$\boldsymbol{\nu} = \left[\frac{\alpha(L, M)}{M-1}, \dots, \frac{\alpha(L, M)}{M-1}, 1 - \alpha(L, M) \right].$$

In the proof of Proposition 2, we have seen that $V^*(\boldsymbol{\nu}) \geq G(\boldsymbol{\nu}) = \alpha(L, M)L = (1 - \nu_M)L$. Furthermore, DP equation (4) implies that $V^*(\boldsymbol{\nu}) \leq \min_{j \in \Omega} (1 - \nu_j)L = (1 - \nu_M)L$. Therefore, $V^*(\boldsymbol{\nu}) = (1 - \nu_M)L$ and the proof is complete.

Remark 9. The technique used to prove Proposition 2 and Corollary 1 can be applied to further improve our proposed policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ by providing a better heuristic for the stopping criteria. More precisely, let \mathcal{R}_i denote the collection of all belief vectors at which it is optimal to retire and declare H_i as the true hypothesis. According to Corollary 1 in [50], $\mathcal{R}_i, i \in \Omega$ are convex. Since $\{\boldsymbol{\rho} \in \mathbb{P}(\Theta) : (1 - \rho_i)L > 1\} \subset \mathcal{R}_i$ (see Remark 1) and $\{\boldsymbol{\rho} \in \mathbb{P}(\Theta) : \rho_i = 1 - \alpha(L, M), \rho_j = \frac{\alpha(L, M)}{M-1} \text{ } j \neq i\} \subset \mathcal{R}_i$ (proved in this appendix), \mathcal{R}_i can be estimated by the convex hull of the above sets.

APPENDIX III

PROOF OF PROPOSITIONS 3 AND 4

A. Proof of Proposition 3

Recall that $\rho_i(n)$ denotes the posterior belief about hypothesis H_i after n observations. Let $\tau, \tau_i, i \in \Omega$, be Markov stopping times defined as follows:

$$\tau := \min \left\{ n : \min_{j \in \Omega} \{1 - \rho_j(n)\} \leq L^{-1} \right\}, \quad (26)$$

$$\tau_i := \min \left\{ n : \min_{j \neq i} \frac{\rho_i(n)}{\rho_j(n)} \geq \frac{1 - L^{-1}}{L^{-1}/(M-1)} \right\}. \quad (27)$$

From (1), total cost under policy $\tilde{\pi}_1$ can be written as

$$\begin{aligned} V^{\tilde{\pi}_1}(\boldsymbol{\rho}) &= \mathbb{E}^{\tilde{\pi}_1}[\tau + \min_{j \in \Omega} (1 - \rho_j(\tau))L] \\ &\leq \mathbb{E}^{\tilde{\pi}_1}[\tau] + 1 \\ &\leq \sum_{i=1}^M \rho_i \mathbb{E}^{\tilde{\pi}_1}[\tau_i | \theta = i] + 1, \end{aligned} \quad (28)$$

where $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_M] = [\rho_1(0), \rho_2(0), \dots, \rho_M(0)]$ and the last inequality follows from the fact that $\tau \leq \tau_i, \forall i \in \Omega$. For notational simplicity, superscript $\tilde{\pi}_1$ is dropped for the rest of the proof.

Next we find an upper bound for $\mathbb{E}[\tau_i|\theta = i]$, $i \in \Omega$. Before we proceed, however, we introduce the following notations to facilitate the proof. Let

$$T^* := \log \frac{1 - L^{-1}}{L^{-1}} - \log \frac{\tilde{\rho}}{1 - \tilde{\rho}},$$

$$\tilde{T}_i(\boldsymbol{\rho}) := \log \frac{\tilde{\rho}}{(1 - \tilde{\rho})/(M - 1)} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k}.$$

For any $\iota > 0$, we have

$$\begin{aligned} & \mathbb{E}[\tau_i|\theta = i] \\ &= \sum_{n=0}^{\infty} P(\{\tau_i > n\}|\theta = i) \\ &\leq 1 + \left(\frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + \frac{T^*}{D_{\mu_i}} \right) (1 + \iota) + \sum_{n: n > \left(\frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + \frac{T^*}{D_{\mu_i}} \right) (1 + \iota)} P(\{\tau_i > n\}|\theta = i) \\ &\stackrel{(a)}{\leq} \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + \frac{T^*}{D_{\mu_i}} + o\left(\frac{\log M}{I_1^2} + \frac{\log L}{I_1 D_{\mu_i}} + M\right) \\ &\leq \frac{\log \frac{\tilde{\rho}}{(1 - \tilde{\rho})/(M - 1)} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k}}{I_1} + \frac{\log \frac{1 - L^{-1}}{L^{-1}} - \log \frac{\tilde{\rho}}{1 - \tilde{\rho}}}{D_{\mu_i}} + \frac{o(M + \log L)}{I_1^2} \\ &= \frac{\log(M - 1)}{I_1} + \left(\log \frac{\tilde{\rho}}{(1 - \tilde{\rho})} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k} \right) \left(\frac{1}{I_1} - \frac{1}{D_{\mu_i}} \right) + \frac{\log \frac{1 - L^{-1}}{L^{-1}} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k}}{D_{\mu_i}} + \frac{o(M + \log L)}{I_1^2} \\ &\leq \frac{\log(M - 1)}{I_1} + \left[\frac{\log \frac{\tilde{\rho}}{(1 - \tilde{\rho})} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k}}{I_1} \right]^+ + \frac{\log \frac{1 - L^{-1}}{L^{-1}} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k}}{D_{\mu_i}} + \frac{o(M + \log L)}{I_1^2} \\ &\leq \frac{\log(M - 1) + \log \frac{1}{\rho_i}}{I_1} + \frac{\log \frac{1 - L^{-1}}{L^{-1}} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k}}{D_{\mu_i}} + \frac{o(M + \log L)}{I_1^2}, \end{aligned} \tag{29}$$

where inequality (a) follows from Lemma 5 below and by setting $\iota = \frac{1}{I_1}(\log ML)^{-\frac{1}{4}}$. Combining (28) and (29) completes the proof of Proposition 3.

Lemma 5. *Given any $\iota > 0$ and for $n > \left(\frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + \frac{T^*}{D_{\mu_i}} \right) (1 + \iota)$, there exists $B(\iota)$ and $b(\iota)$ as shown in (33) such that $P(\{\tau_i > n\}|\theta = i) \leq MB(\iota)e^{-b(\iota)n}$.*

Proof of Lemma 5:

Let $B_{ij}(n)$ and $\tilde{B}_{ij}(n)$ be events in the probability space defined as follows:

$$B_{ij}(n) := \left\{ \log \frac{\rho_i(n)}{\rho_j(n)} < \log \frac{1 - L^{-1}}{L^{-1}/(M - 1)} \right\},$$

$$\tilde{B}_{ij}(n) := \left\{ \log \frac{\rho_i(n)}{\rho_j(n)} < \log \frac{\tilde{\rho}}{(1 - \tilde{\rho})/(M - 1)} \right\}.$$

We have,

$$\begin{aligned}
& P(\tilde{B}_{ij}(n)|\theta = i) \\
&= P\left(\left\{\log \frac{\rho_i(n)}{\rho_j(n)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] < \log \frac{\tilde{\rho}}{(1 - \tilde{\rho})/(M - 1)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}]\right\}|\theta = i\right) \\
&= P\left(\left\{\log \frac{\rho_i(n)}{\rho_j(n)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] < \log \frac{\tilde{\rho}}{(1 - \tilde{\rho})/(M - 1)} - \mathbb{E}\left[\log \frac{\rho_i}{\rho_j} + \sum_{t=0}^{n-1} \log \frac{q_i^{A(t)}}{q_j^{A(t)}}\right]\right\}|\theta = i\right) \\
&\leq P\left(\left\{\log \frac{\rho_i(n)}{\rho_j(n)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] < \log \frac{\tilde{\rho}}{(1 - \tilde{\rho})/(M - 1)} - \min_{k \neq i} \log \frac{\rho_i}{\rho_k} - nI_1\right\}|\theta = i\right) \\
&= P\left(\left\{\log \frac{\rho_i(n)}{\rho_j(n)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] < \tilde{T}_i(\boldsymbol{\rho}) - nI_1\right\}|\theta = i\right). \tag{30}
\end{aligned}$$

Similarly, we can show that

$$P(B_{ij}(n)|\theta = i) \leq P\left(\left\{\log \frac{\rho_i(n)}{\rho_j(n)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] < T^* - (n - \tilde{\tau}_i)D_{\mu_i}\right\}|\theta = i\right), \tag{31}$$

where

$$\tilde{\tau}_i = \min \left\{ n : \min_{j \neq i} \frac{\rho_i(n')}{\rho_j(n')} \geq \frac{\tilde{\rho}}{(1 - \tilde{\rho})/(M - 1)} \quad \forall n' \geq n \right\}.$$

By construction (27),

$$\begin{aligned}
& P(\{\tau_i > n\}|\theta = i) \\
&\leq P(\cup_{j \neq i} B_{ij}(n)|\theta = i) \\
&\leq P(\cup_{j \neq i} B_{ij}(n) \cap \{\tilde{\tau}_i \leq \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + n\frac{\iota/2}{1+\iota}\}|\theta = i) + P(\{\tilde{\tau}_i > \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + n\frac{\iota/2}{1+\iota}\}|\theta = i) \\
&\leq \sum_{j \neq i} \left(P(B_{ij}(n) \cap \{\tilde{\tau}_i \leq \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + n\frac{\iota/2}{1+\iota}\}|\theta = i) + \sum_{m: m > \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + n\frac{\iota/2}{1+\iota}} P(\tilde{B}_{ij}(m)|\theta = i) \right) \\
&= \sum_{j \neq i} \left(P\left(\left\{\log \frac{\rho_i(n)}{\rho_j(n)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] < D_{\mu_i}(\frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + \frac{T^*}{D_{\mu_i}} - n\frac{1+\iota/2}{1+\iota})\right\}|\theta = i\right) \right. \\
&\quad \left. + \sum_{m: m > \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + n\frac{\iota/2}{1+\iota}} P\left(\left\{\log \frac{\rho_i(m)}{\rho_j(m)} - \mathbb{E}[\log \frac{\rho_i(m)}{\rho_j(m)}] < \tilde{T}_i(\boldsymbol{\rho}) - mI_1\right\}|\theta = i\right) \right). \tag{32}
\end{aligned}$$

For any $a, \hat{a} \in \mathcal{A}$ and $i, j \in \Omega$, we have $\left| \log \frac{q_i^a}{q_j^a} - \log \frac{q_i^{\hat{a}}}{q_j^{\hat{a}}} \right| \leq 2 \log \xi$. For $k = 1, 2, \dots, n$, let $X_k = \log \frac{q_i^{A(k-1)}}{q_j^{A(k-1)}}$ and $\mathbf{X} = [X_1, X_2, \dots, X_n]$. Define function $f(\mathbf{X}) = \log \frac{\rho_i}{\rho_j} + \sum_{k=1}^n X_k = \log \frac{\rho_i(n)}{\rho_j(n)}$. From (32) and Fact 3 below, and for $n > \left(\frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + \frac{T^*}{D_{\mu_i}} \right) (1 + \iota)$, we have

$$\begin{aligned}
& P(\{\tau_i > n\} | \theta = i) \\
& \leq (M-1) \left[\exp \left(-2n \left(\frac{D_{\mu_i}/(1+\iota)}{2 \log \xi} \right)^2 \left(1 + \iota/2 - \frac{1}{n} \left(\frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + \frac{T^*}{D_{\mu_i}} \right) (1+\iota) \right)^2 \right) \right. \\
& \quad \left. + \sum_{m: m > \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + n \frac{\iota/2}{1+\iota}} \exp \left(-2m \left(\frac{I_1/(1+\iota/2)}{2 \log \xi} \right)^2 \left(1 + \iota/2 - \frac{1}{m} \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} (1+\iota/2) \right)^2 \right) \right] \\
& \leq (M-1) \left[\exp \left(-n \frac{\iota^2}{2} \left(\frac{D_{\mu_i}/(1+\iota)}{2 \log \xi} \right)^2 \right) + \sum_{m: m > \frac{\tilde{T}_i(\boldsymbol{\rho})}{I_1} + n \frac{\iota/2}{1+\iota}} \exp \left(-m \frac{\iota^2}{2} \left(\frac{I_1/(1+\iota/2)}{2 \log \xi} \right)^2 \right) \right] \\
& \leq (M-1) \left[\exp \left(-n \frac{\iota^2}{2} \left(\frac{D_{\mu_i}/(1+\iota)}{2 \log \xi} \right)^2 \right) + \frac{\exp \left(-n \frac{\iota^3}{4(1+\iota)} \left(\frac{I_1/(1+\iota/2)}{2 \log \xi} \right)^2 \right)}{1 - \exp \left(-\frac{\iota^2}{2} \left(\frac{I_1/(1+\iota/2)}{2 \log \xi} \right)^2 \right)} \right]. \tag{33}
\end{aligned}$$

■

Fact 3 (McDiarmid's Inequality [51]). *Let $\mathbf{X} = [X_1, X_2, \dots, X_n]$ be a family of independent random variables with X_k taking values in a set \mathcal{X}_k for each k . Suppose a real-valued function f defined on $\prod_{k=1}^n \mathcal{X}_k$ satisfies $|f(\mathbf{x}) - f(\mathbf{x}')| \leq c_k$, whenever the vectors \mathbf{x} and \mathbf{x}' only differ in the k -th coordinate. Then for any $\nu > 0$,*

$$P(f(\mathbf{X}) - \mathbb{E}[f(\mathbf{X})] \leq -\nu) \leq e^{-2\nu^2 / \sum_{k=1}^n c_k^2}.$$

B. Proof of Proposition 4

Recall that $\rho_i(n)$ denotes the posterior belief about hypothesis H_i after n observations. Let τ , τ_i , $i \in \Omega$, be Markov stopping times defined as follows:

$$\tau := \min \left\{ n : \min_{j \in \Omega} \{1 - \rho_j(n)\} \leq L^{-1} \right\}, \tag{34}$$

$$\tau_i := \min \left\{ n : \rho_i(n) \geq 1 - L^{-1} \right\}. \tag{35}$$

From (1), total cost under policy $\tilde{\pi}_2$ is upper bounded as

$$\begin{aligned}
V^{\tilde{\pi}_2}(\boldsymbol{\rho}) &= \mathbb{E}^{\tilde{\pi}_2}[\tau + \min_{j \in \Omega} (1 - \rho_j(\tau))L] \\
&\leq \mathbb{E}^{\tilde{\pi}_2}[\tau] + 1 \\
&\leq \sum_{i=1}^M \rho_i \mathbb{E}^{\tilde{\pi}_2}[\tau_i | \theta = i] + 1, \tag{36}
\end{aligned}$$

where $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_M] = [\rho_1(0), \rho_2(0), \dots, \rho_M(0)]$ and the last inequality follows from the fact that $\tau \leq \tau_i, \forall i \in \Omega$. For notational simplicity, superscript $\tilde{\pi}_2$ is dropped for the rest of the proof.

Next we find an upper bound for $\mathbb{E}[\tau_i | \theta = i], i \in \Omega$. Let $U_n := \log \frac{\rho_i(n)}{1-\rho_i(n)} - \log \frac{\tilde{\rho}}{1-\tilde{\rho}}$ and let \mathcal{F}_n denote the history of previous actions and observations up to time n , i.e., $\mathcal{F}_n = \sigma\{\boldsymbol{\rho}(0), A(0), Z(0), \dots, A(n-1), Z(n-1)\}$. Under policy $\tilde{\pi}_2$, the sequence $\{U_n\}, n = 0, 1, \dots$, forms a submartingale with respect to the filtration $\{\mathcal{F}_n\}$ with the following properties:

$$(C1) \quad |U_n - U_{n-1}| \leq \max_{i,j \in \Omega} \max_{a \in \mathcal{A}} \sup_{z \in \mathcal{Z}} \log \frac{q_i^a(z)}{q_j^a(z)} \leq \log \xi,$$

$$(C2) \quad \text{If } U_n \geq 0 \text{ (} \rho_i(n) \geq \tilde{\rho} \Rightarrow P(\{A(n) = a\}) = \eta_{ia}\text{):}$$

$$\begin{aligned} & \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i] \\ &= \sum_{a \in \mathcal{A}} P(\{A(n) = a\}) \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i, A(n) = a] \\ &= \sum_{a \in \mathcal{A}} \eta_{ia} \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i, A(n) = a] \\ &= \sum_{a \in \mathcal{A}} \eta_{ia} \mathbb{E} \left[\log \frac{\rho_i(n) q_i^a(Z)}{\sum_{j \neq i} \rho_j(n) q_j^a(Z)} - \log \frac{\rho_i(n)}{1 - \rho_i(n)} \middle| \mathcal{F}_n, \theta = i \right] \\ &= \sum_{a \in \mathcal{A}} \eta_{ia} \int q_i^a(z) \log \frac{q_i^a(z)}{\sum_{j \neq i} \frac{\rho_j(n)}{1 - \rho_i(n)} q_j^a(z)} dz \\ &\geq \max_{\boldsymbol{\lambda} \in \Lambda(\mathcal{A})} \min_{\hat{\boldsymbol{\rho}} \in \mathbb{P}_L(\Theta)} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a) \\ &= D_{\boldsymbol{\eta}_i}, \end{aligned}$$

$$(C3) \quad \text{If } U_n < 0 \text{ and } \rho_j(n) < \tilde{\rho} \text{ for all } j \text{ (} \Rightarrow P(\{A(n) = a\}) = \eta_{0a}\text{):}$$

$$\begin{aligned} & \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i] \\ &= \sum_{a \in \mathcal{A}} \eta_{0a} \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i, A(n) = a] \\ &= \sum_{a \in \mathcal{A}} \eta_{0a} \int q_i^a(z) \log \frac{q_i^a(z)}{\sum_{j \neq i} \frac{\rho_j(n)}{1 - \rho_i(n)} q_j^a(z)} dz \\ &\geq \max_{\boldsymbol{\lambda} \in \Lambda(\mathcal{A})} \min_{i \in \Omega} \min_{\hat{\boldsymbol{\rho}} \in \mathbb{P}_L(\Theta)} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a) \\ &= I_2. \end{aligned}$$

Stopping time τ_i can be rewritten as

$$\begin{aligned}
\tau_i &= \min \left\{ n : \rho_i(n) \geq 1 - L^{-1} \right\} \\
&= \min \left\{ n : \frac{\rho_i(n)}{1 - \rho_i(n)} \geq \frac{1 - L^{-1}}{L^{-1}} \right\} \\
&= \min \left\{ n : \log \frac{\rho_i(n)}{1 - \rho_i(n)} - \log \frac{\tilde{\rho}}{1 - \tilde{\rho}} \geq \log \frac{1 - L^{-1}}{L^{-1}} - \log \frac{\tilde{\rho}}{1 - \tilde{\rho}} \right\}. \tag{37}
\end{aligned}$$

The assertion of the proposition follows from (37) and the following lemma which is a slight generalization of Lemma 6 in [11].

Lemma 6. *Assume that the sequence $\{\zeta_n\}$, $n = 0, 1, \dots$ forms a submartingale with respect to a filtration $\{\mathcal{F}_n\}$. Furthermore, assume there exist positive constants $K_1 \leq K_2 \leq K_3$ such that*

$$\begin{aligned}
\mathbb{E}[\zeta_{n+1} | \mathcal{F}_n] &\geq \zeta_n + K_1 \quad \text{if } \zeta_n < 0, \\
\mathbb{E}[\zeta_{n+1} | \mathcal{F}_n] &\geq \zeta_n + K_2 \quad \text{if } \zeta_n \geq 0, \\
|\zeta_{n+1} - \zeta_n| &\leq K_3.
\end{aligned}$$

Consider the stopping time $\tau_B = \min\{n : \zeta_n \geq B\}$, $B > 0$. Then we have the inequality

$$\mathbb{E}[\tau_B] \leq \frac{B - \zeta_0}{K_2} + \zeta_0 \mathbf{1}_{\{\zeta_0 < 0\}} \left(\frac{1}{K_2} - \frac{1}{K_1} \right) + 3 \frac{K_3^2}{K_1 K_2}.$$

In particular, from (C1)–(C3) and Lemma 6, we have

$$\begin{aligned}
\rho_i \mathbb{E}[\tau_i | \theta = i] &\leq \rho_i \left(\frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\tilde{\rho}}{1-\tilde{\rho}}}{D_{\eta_i}} + \frac{-\left(\log \frac{\rho_i}{1-\rho_i} - \log \frac{\tilde{\rho}}{1-\tilde{\rho}}\right)}{I_2} + 3 \frac{(\log \xi)^2}{I_2 D_{\eta_i}} \right) \\
&= \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}}}{D_{\eta_i}} + \frac{\rho_i \log \frac{1-\rho_i}{\rho_i}}{I_2} + \rho_i \frac{\log \frac{\tilde{\rho}}{1-\tilde{\rho}} (D_{\eta_i} - I_2) + 3(\log \xi)^2}{I_2 D_{\eta_i}} \\
&\leq \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}}}{D_{\eta_i}} + \frac{\rho_i \log \frac{1-\rho_i}{\rho_i}}{I_2} + \rho_i \frac{K_2''}{I_2 D_{\eta_i}}, \tag{38}
\end{aligned}$$

where $K_2'' = \log \frac{\tilde{\rho}}{1-\tilde{\rho}} \overline{D}_{max} + 3(\log \xi)^2$ is independent of L and M . Now from (36), (38), and the fact that $\sum_{i=1}^M \rho_i \log \frac{1-\rho_i}{\rho_i} < H(\rho)$, we have the assertion of the proposition.

APPENDIX IV

PROOF OF PROPOSITION 5

V_λ is the solution to the following DP equation:

$$V_\lambda(\rho) = \min \left\{ 1 + \sum_{a \in \mathcal{A}} \lambda_a (\mathbb{T}^a V_\lambda)(\rho), \min_{j \in \Omega} (1 - \rho_j) L \right\}.$$

Let Γ be the set of all mappings $\gamma : \Omega \rightarrow \Omega$ such that $\gamma(i) \neq i$ for $i \in \Omega$. Now associated with any $\gamma \in \Gamma$ define

$$\underline{V}_\lambda^\gamma(\boldsymbol{\rho}) = \left[\sum_{i=1}^M \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i}{\rho_{\gamma(i)}}}{\sum_{\hat{a} \in \mathcal{A}} \lambda_{\hat{a}} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} - K'_\lambda \right]^+.$$

Next we use a slight modification of Lemma 1 in which $\min_{a \in \mathcal{A}} (\mathbb{T}^a V)(\boldsymbol{\rho})$ is replaced by $\sum_{a \in \mathcal{A}} \lambda_a (\mathbb{T}^a V)(\boldsymbol{\rho})$ to show that $V_\lambda \geq \underline{V}_\lambda^\gamma$ for all $\gamma \in \Gamma$. In particular, we show that for all $\gamma \in \Gamma$ and all $\boldsymbol{\rho} \in \mathbb{P}(\Theta)$, $\underline{V}_\lambda^\gamma(\boldsymbol{\rho}) \leq \min \{1 + \sum_{a \in \mathcal{A}} \lambda_a (\mathbb{T}^a \underline{V}_\lambda^\gamma)(\boldsymbol{\rho}), \min_{j \in \Omega} (1 - \rho_j)L\}$. For any $\boldsymbol{\rho}$ such that $\underline{V}_\lambda^\gamma(\boldsymbol{\rho}) = 0$, the inequality holds trivially. For $\underline{V}_\lambda^\gamma(\boldsymbol{\rho}) > 0$ and for any action $a \in \mathcal{A}$ we have

$$\begin{aligned} & \sum_{a \in \mathcal{A}} \lambda_a (\mathbb{T}^a \underline{V}_\lambda^\gamma)(\boldsymbol{\rho}) \\ & \geq \sum_{a \in \mathcal{A}} \lambda_a \left(\sum_{i=1}^M \int \rho_i q_i^a(z) \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i q_i^a(z)}{\rho_{\gamma(i)} q_{\gamma(i)}^a(z)}}{\sum_{\hat{a} \in \mathcal{A}} \lambda_{\hat{a}} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} dz - K'_\lambda \right) \\ & \geq \underline{V}_\lambda^\gamma(\boldsymbol{\rho}) - \sum_{i=1}^M \rho_i \sum_{a \in \mathcal{A}} \lambda_a \int q_i^a(z) \frac{\log \frac{q_i^a(z)}{q_{\gamma(i)}^a(z)}}{\sum_{\hat{a} \in \mathcal{A}} \lambda_{\hat{a}} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} dz \\ & = \underline{V}_\lambda^\gamma(\boldsymbol{\rho}) - 1. \end{aligned}$$

Claim 5 (in Appendix VII). Constant K'_λ can be selected independent of L such that $\underline{V}_\lambda^\gamma(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j)L$ is satisfied for all $\gamma \in \Gamma$.

Using Claim 5 and letting $\underline{V}_\lambda(\cdot) = \max_{\gamma \in \Gamma} \underline{V}_\lambda^\gamma(\cdot)$, we have the assertion of the proposition.

APPENDIX V

INFORMATION ACQUISITION RATE AND RELIABILITY

A. Proof of Lemma 2

Let $\tau_L := \min\{n : \max_{j \in \Omega} \rho_j(n) \geq 1 - L^{-1}\}$. Then

$$\begin{aligned} \mathbb{E}[\tau_\epsilon^*] & \geq \mathbb{E}[\tau_\epsilon^* | \max_{j \in \Omega} \rho_j(\tau_\epsilon^*) \geq 1 - L^{-1}] P(\{\max_{j \in \Omega} \rho_j(\tau_\epsilon^*) \geq 1 - L^{-1}\}) \\ & \stackrel{(a)}{\geq} \mathbb{E}[\tau_\epsilon^* | \max_{j \in \Omega} \rho_j(\tau_\epsilon^*) \geq 1 - L^{-1}] (1 - \mathbb{E}[1 - \max_{j \in \Omega} \rho_j(\tau_\epsilon^*)]L) \\ & \stackrel{(b)}{\geq} \mathbb{E}[\tau_\epsilon^* | \max_{j \in \Omega} \rho_j(\tau_\epsilon^*) \geq 1 - L^{-1}] (1 - \epsilon L) \\ & \geq \mathbb{E}[\tau_L] (1 - \epsilon L), \end{aligned} \tag{39}$$

where (a) follows from Markov inequality and (b) follows from the definition of τ_ϵ^* which implies that $\text{Pe} = \mathbb{E}[1 - \max_{j \in \Omega} \rho_j(\tau_\epsilon^*)] \leq \epsilon$.

Let $V_L : \mathbb{P}(\Theta) \rightarrow \mathbb{R}_+$ be the minimal solution to the following fixed point equation:

$$V_L(\boldsymbol{\rho}) = \min \left\{ 1 + \min_{a \in \mathcal{A}} \{(\mathbb{T}^a V_L)(\boldsymbol{\rho})\}, \mathfrak{L}(\boldsymbol{\rho}) \right\}, \quad (40)$$

where

$$\mathfrak{L}(\boldsymbol{\rho}) = \begin{cases} 0 & \text{if } \min_{j \in \Omega} (1 - \rho_j) \leq L^{-1} \\ \infty & \text{otherwise} \end{cases}. \quad (41)$$

It can be easily shown that

$$\mathbb{E}[\tau_L] = V_L(\boldsymbol{\rho}(0)) \geq V^*(\boldsymbol{\rho}(0)) - 1. \quad (42)$$

Combining (39) and (42) completes the proof.

B. Proof of Corollary 3

Set $\epsilon = 2^{-Et}$, $L = \frac{1}{\epsilon \log \frac{1}{\epsilon}}$, and $\delta = \frac{1}{\log \frac{1}{\epsilon}}$. For $L > \frac{\log M}{I_{max}}$ and $t > \frac{4}{E}$, we can use Lemma 2 and Proposition 2 to obtain the following lower bound

$$\mathbb{E}[\tau_\epsilon^*] \geq (1 - \frac{1}{Et}) \left[\frac{H(\boldsymbol{\rho}(0)) - \frac{1}{Et} \log M}{I_{max}} + \frac{\log \frac{2^{Et}}{Et} - \log Et}{D_{max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i(0) < 1 - \frac{1}{Et}\}} - O(1) \right]^+.$$

For uniform prior $\boldsymbol{\rho}(0) = [1/M, \dots, 1/M]$, the lower bound simplifies to

$$\mathbb{E}[\tau_\epsilon^*] \geq \left[\frac{(1 - \frac{1}{Et})^2 \log M}{I_{max}} + (1 - \frac{1}{Et}) \frac{Et - 2 \log Et}{D_{max}} - O(1) \right]^+.$$

Therefore, for any policy π ,

$$\frac{(1 - \frac{1}{Et})^2 \log M^\pi(t, 2^{-Et})}{\bar{I}_{max}} + (1 - \frac{1}{Et}) \frac{Et - 2 \log Et}{\bar{D}_{max}} - O(1) \leq t,$$

and hence,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \log M^\pi(t, 2^{-Et}) &\leq \lim_{t \rightarrow \infty} \frac{1}{t} \frac{t - (1 - \frac{1}{Et}) \frac{Et - 2 \log Et}{\bar{D}_{max}} + O(1)}{(1 - \frac{1}{Et})^2} \bar{I}_{max} \\ &= \lim_{t \rightarrow \infty} \frac{1 - \frac{E}{\bar{D}_{max}} + \frac{2(1 - \frac{1}{Et}) \log Et + 1}{t \bar{D}_{max}} + O(\frac{1}{t})}{(1 - \frac{1}{Et})^2} \bar{I}_{max} \\ &= \bar{I}_{max} \left(1 - \frac{E}{\bar{D}_{max}} \right) \end{aligned} \quad (43)$$

$$< \bar{I}_{max}. \quad (44)$$

Inequality (44) implies that for $E > 0$, no policy can achieve rates greater than \bar{I}_{max} . Furthermore, using (43), the reliability function can be bounded as

$$E(R) \leq \bar{D}_{max} \left(1 - \frac{R}{\bar{I}_{max}} \right).$$

Next we show that for fixed M , hence at $R = 0$, no policy can achieve reliability higher than D_1 . From Lemma 2 and Fact 2, and for uniform prior $\boldsymbol{\rho}(0) = [1/M, \dots, 1/M]$, we obtain the following lower bound

$$\mathbb{E}[\tau_\epsilon^*] \geq (1 - \epsilon L) \left(\frac{\log L}{D_1} - o(\log L) \right).$$

Using the inequality above, we can find a lower bound on ϵ such that $\mathbb{E}[\tau_\epsilon^*] \leq t$ is satisfied. More precisely, for any policy π we obtain

$$\text{Pe}^\pi(t, M) \geq L^{-1} \left(1 - \frac{t}{\frac{\log L}{D_1} - o(\log L)} \right).$$

We can select $L = 2^{D_1 t + o(t)}$ such that it satisfies

$$\frac{1}{\text{Pe}^\pi(t, M)} \leq 2^{D_1 t + o(t)} O(t), \quad (45)$$

and hence,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{-1}{t} \log \text{Pe}^\pi(t, M) &\leq \lim_{t \rightarrow \infty} \frac{1}{t} \log (2^{D_1 t + o(t)} O(t)) \\ &= \lim_{t \rightarrow \infty} \left(D_1 + \frac{o(t)}{t} + O\left(\frac{\log t}{t}\right) \right) \\ &= D_1. \end{aligned} \quad (46)$$

C. Generalizing the Result of Corollary 3

The result of Corollary 3 can be strengthened to show that no policy can achieve diminishing error probability at rates higher than \bar{I}_{max} . Here we provide the sketch of the proof.

Using (40) and a slight modification of Lemma 1 in which $\min_{j \in \Omega} (1 - \rho_j) L$ is replaced by $\mathfrak{L}(\boldsymbol{\rho})$ (as defined in (41)), we can find the following lower bound for $V_L(\boldsymbol{\rho})$:

$$\underline{V}_L(\boldsymbol{\rho}) := \left[\frac{H(\boldsymbol{\rho}) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{max}} + \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - \hat{K} \right]^+, \quad (47)$$

where $\hat{K} = \max \left\{ \frac{L^{-1} \log(M-1)+1}{I_{max}}, \frac{L^{-1} \log(M-1)+\log \xi+2}{D_{max}} \right\} \leq \frac{L^{-1} \log(M-1)+1}{I_{max}} + \frac{\log \xi+1}{D_{max}}.$

Combining (39), (42), and (47), we get

$$\mathbb{E}[\tau_\epsilon^*] \geq (1 - \epsilon L) \underline{V}_L(\boldsymbol{\rho}(0)).$$

Let $u(t)$ be a function such that $u(t) \rightarrow 0$ as $t \rightarrow \infty$ but for any $E > 0$, $u(t)2^{Et} \rightarrow \infty$. In other words, $\frac{\log \frac{1}{u(t)}}{t} \rightarrow 0$ as $t \rightarrow \infty$. Set $\epsilon = u(t)$, $L = \frac{1}{\sqrt{u(t)}}$, and $\delta = \sqrt{u(t)}$. We obtain the following lower bound

$$\mathbb{E}[\tau_\epsilon^*] \geq (1 - \sqrt{u(t)}) \left[\frac{\log M - 2\sqrt{u(t)} \log M}{I_{max}} - O(1) \right]^+. \quad (48)$$

Therefore, for any policy π ,

$$\frac{(1 - 3\sqrt{u(t)} + 2u(t)) \log M^\pi(t, u(t))}{\bar{I}_{max}} - O(1) \leq t,$$

and hence,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log M^\pi(t, u(t)) \leq \lim_{t \rightarrow \infty} \frac{1}{t} \frac{t + O(1)}{1 - 3\sqrt{u(t)} + 2u(t)} \bar{I}_{max} = \bar{I}_{max}. \quad (49)$$

D. Proof of Corollary 4

By definition, policy π is asymptotically optimal in L if $V^\pi(\rho) = \mathbb{E}^\pi[\tau] + L\text{Pe}^\pi \leq \frac{\log L}{D_1} + o(\log L)$. This implies that policy π can achieve $\text{Pe}^\pi \leq L^{-1} \left(\frac{\log L}{D_1} + o(\log L) \right)$ with $\mathbb{E}^\pi[\tau] \leq \frac{\log L}{D_1} + o(\log L)$. To satisfy $\mathbb{E}^\pi[\tau] \leq t$, L can be selected as $L = 2^{D_1 t(1-o(1))}$ where $o(1) \rightarrow 0$ as $t \rightarrow \infty$. For this selection of L , $\text{Pe}^\pi(t, M)$ is bounded as

$$\text{Pe}^\pi(t, M) \leq 2^{-D_1 t(1-o(1))} t,$$

and by definition, the error exponent of policy π is

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{-1}{t} \log \text{Pe}^\pi(t, M) &\geq \lim_{t \rightarrow \infty} \frac{-1}{t} (\log 2^{-D_1 t(1-o(1))} + \log t) \\ &= \lim_{t \rightarrow \infty} \left(D_1(1 - o(1)) - \frac{\log t}{t} \right) \\ &= D_1. \end{aligned} \quad (50)$$

E. Proof of Corollary 5

We show that if policy π is not an order optimal solution to Problem (P), then it cannot achieve non-zero rate with non-zero reliability.

Let $\mathbb{E}^\pi[\tau_\epsilon]$ denote the expected number of samples that policy π requires to achieve $\text{Pe}^\pi \leq \epsilon$. If policy π is not order optimal in L and M , then $\mathbb{E}^\pi[\tau_\epsilon]$ is either 1) $\omega(\log M) + O(\log \frac{1}{\epsilon})$ where $\frac{\omega(\log M)}{\log M} \rightarrow \infty$ as $M \rightarrow \infty$; 2) $O(\log M) + \omega(\log \frac{1}{\epsilon})$; or 3) $\omega(\log M) + \omega(\log \frac{1}{\epsilon})$. Proof is done by contradiction. Suppose $\mathbb{E}^\pi[\tau_\epsilon] = O(\log M) + O(\log \frac{1}{\epsilon})$. Then

$$V^\pi(\rho) = \mathbb{E}^\pi[\tau] + L\text{Pe}^\pi \leq \mathbb{E}^\pi[\tau_\epsilon] + L\epsilon|_{\epsilon=L^{-1}} = O(\log M) + O(\log L),$$

which is an order optimal solution to Problem (P).

- **Case 1:** $\mathbb{E}^\pi[\tau_\epsilon] = \omega(\log M) + O(\log \frac{1}{\epsilon})$.

Setting $\epsilon = 2^{-Et}$ for some $E > 0$, we obtain from condition above that $\log M^\pi(t, 2^{-Et}) = o(t)$.

By definition,

$$\begin{aligned} R &= \lim_{t \rightarrow \infty} \frac{1}{t} \log M^\pi(t, 2^{-Et}) \\ &= \lim_{t \rightarrow \infty} \frac{o(t)}{t} = 0. \end{aligned}$$

- **Case 2:** $\mathbb{E}^\pi[\tau_\epsilon] = O(\log M) + \omega(\log \frac{1}{\epsilon})$.

Setting $M = 2^{Rt}$ for some $R > 0$, we obtain from condition above that $-\log \text{Pe}^\pi(t, 2^{Rt}) = o(t)$.

By definition,

$$\begin{aligned} E &= \lim_{t \rightarrow \infty} \frac{-1}{t} \log \text{Pe}^\pi(t, 2^{Rt}) \\ &= \lim_{t \rightarrow \infty} \frac{o(t)}{t} = 0. \end{aligned}$$

- **Case 3:** $\mathbb{E}^\pi[\tau_\epsilon] = \omega(\log M) + \omega(\log \frac{1}{\epsilon})$.

Proof follows similar lines as the proof of Case 1 and 2.

F. Proof of Corollary 6

As shown in the proof of \bar{V}_2 in Appendix III, policy $\tilde{\pi}_2$ can achieve $\text{Pe}^{\tilde{\pi}_2} \leq L^{-1}$ with $\mathbb{E}^{\tilde{\pi}_2}[\tau] \leq \frac{\log M}{I_2} + \frac{\log L}{D_2} + \frac{O(1)}{I_2 D_2}$. For $L = 2^{Et}$ we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \log M^{\tilde{\pi}_2}(t, 2^{-Et}) &\geq \lim_{t \rightarrow \infty} \frac{t - \frac{Et}{D_2} - \frac{O(1)}{I_2 D_2}}{t} I_2 \\ &= I_2 \left(1 - \frac{E}{D_2} \right). \end{aligned} \quad (51)$$

Therefore, for any rate $R \in [0, I_2]$, there exists a reliability E , $E \geq \underline{D}_2 \left(1 - \frac{R}{I_2} \right)$, such that $\tilde{\pi}_2$ can achieve rate R with reliability E .

G. Achievable Rate and Reliability Region for π^*

By definition, $V^*(\rho) = \mathbb{E}^{\pi^*}[\tau] + L \text{Pe}^{\pi^*}$. Thus, $\mathbb{E}^{\pi^*}[\tau] \leq V^*(\rho) \leq \bar{V}_2(\rho)$ and $\text{Pe}^{\pi^*} \leq L^{-1} V^*(\rho) \leq L^{-1} \bar{V}_2(\rho)$. Let $L = t 2^{Et}$ and suppose the hypotheses are equiprobable. We obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \log M^{\pi^*}(t, 2^{-Et}) &\geq \lim_{t \rightarrow \infty} \frac{t - \frac{\log t 2^{Et}}{D_2} - \frac{O(1)}{I_2 D_2} - 1}{t} I_2 \\ &= I_2 \left(1 - \frac{E}{D_2} \right). \end{aligned} \quad (52)$$

Therefore, for any rate $R \in [0, I_2]$, there exists a reliability E , $E \geq \underline{D}_2 \left(1 - \frac{R}{I_2} \right)$, such that π^* can achieve rate R with reliability E .

APPENDIX VI

PROOF OF LEMMA 4

Recall that $I_{max} = \max_{e \in \mathcal{E}} \max_{\hat{\rho} \in \mathbb{P}(\Theta)} I(\hat{\rho}; q_{\hat{\rho}}^e)$ is the maximum (taken over the distribution of θ) mutual information between θ and observation Y . Since $\theta \rightarrow X \rightarrow Y$ forms a Markov chain, we obtain from data processing inequality that $I_{max} \leq \max_{P_X} I(X; Y) = C$, where P_X denotes the distribution of X .

Let P_X^* be the capacity achieving probability distribution defined on \mathcal{X} . Let λ^* be such that for all $e \in \mathcal{E}$, $\lambda_e^* = \Pi_{i=1}^M P_X^*(e(i))$. We have,

$$\begin{aligned}
I_2 &= \max_{\lambda \in \Lambda(\mathcal{E})} \min_{i \in \Omega} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{e \in \mathcal{E}} \lambda_e D(q_i^e \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^e) \\
&\geq \min_{i \in \Omega} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{e \in \mathcal{E}} \lambda_e^* D(q_i^e \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^e) \\
&\stackrel{(a)}{\geq} \min_{i \in \Omega} \sum_{e \in \mathcal{E}} \lambda_e^* D(q_i^e \| \sum_{x' \in \mathcal{X}} P_X^*(x') P(Y|X = x')) \\
&= \sum_{x \in \mathcal{X}} P_X^*(x) D(P(Y|X = x) \| \sum_{x' \in \mathcal{X}} P_X^*(x') P(Y|X = x')) \\
&\stackrel{(b)}{=} C,
\end{aligned}$$

where (a) follows from Jensen's inequality and the fact that under randomization λ^* , observation kernels q_i^e and q_j^e , $i \neq j$, are independent from each other; and (b) follows from Theorem 4.5.1 in [32].

From the discussion above, we obtain $\bar{I}_{max} \leq C \leq \underline{I}_2$. Equality (15) follows from Corollary 3 which implies that no policy can achieve rates higher than \bar{I}_{max} and hence, $\underline{I}_2 \leq \bar{I}_{max}$.

Next we prove equality (16). Let $x, x' \in \mathcal{X}$ be two inputs of the channel satisfying $D(P(Y|X = x) \| P(Y|X = x')) = C_1$. We have,

$$\begin{aligned}
D_{max} &= \max_{i, j \in \Omega} \max_{e \in \mathcal{E}} D(q_i^e \| q_j^e) \\
&= \max_{i, j \in \Omega} \max_{e \in \mathcal{E}} D(P(Y|X = e(i)) \| P(Y|X = e(j))) \\
&= D(P(Y|X = x) \| P(Y|X = x')) \\
&= C_1.
\end{aligned} \tag{53}$$

Let $e_i^* \in \mathcal{E}$, $i \in \Omega$, be such that $e_i^*(i) = x$ and for all $j \neq i$, $e_i^*(j) = x'$. For all $i \in \Omega$, we get

$$\begin{aligned}
\max_{\lambda \in \Lambda(\mathcal{E})} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{e \in \mathcal{E}} \lambda_e D(q_i^e \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^e) &\geq \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} D(q_i^{e_i^*} \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^{e_i^*}) \\
&= D(P(Y|X = x) \| P(Y|X = x')) \\
&= C_1,
\end{aligned}$$

and by definition,

$$D_2 = M \left(\sum_{i=1}^M \frac{1}{\max_{\lambda \in \Lambda(\mathcal{E})} \min_{\hat{\rho} \in \mathbb{P}_L(\Theta)} \sum_{e \in \mathcal{E}} \lambda_e D(q_i^e \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1-\hat{\rho}_i} q_j^e)} \right)^{-1} \geq M \left(\sum_{i=1}^M \frac{1}{C_1} \right)^{-1} = C_1. \quad (54)$$

Combining (53) and (54), we obtain $\bar{D}_{max} = C_1 \leq \underline{D}_2$. Equality (16) follows from Corollary 3 which implies that no policy can achieve reliability higher than \bar{D}_{max} and hence, $\underline{D}_2 \leq \bar{D}_{max}$.

APPENDIX VII

PROOF OF TECHNICAL CLAIMS

A. Proof of Claim 1

Let $D_{min} = \min_{i,j \in \Omega} \max_{a \in \mathcal{A}} D(q_i^a \| q_j^a)$. First we notice that, if $\boldsymbol{\rho} \notin \mathbb{P}_L(\Theta)$, then \underline{V}_1 is bounded as:

$$\begin{aligned} \underline{V}_1(\boldsymbol{\rho}) &\leq \left[\sum_{i=1}^M \rho_i \max_{j \neq i} \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i}{1-\rho_i}}{\max_{a \in \mathcal{A}} D(q_i^a \| q_j^a)} - K'_1 \right]^+ \\ &\leq \left[\sum_{\{i \in \Omega: \rho_i < 1-L^{-1}\}} \rho_i \frac{\log L + \log \frac{1}{\rho_i}}{D_{min}} - K'_1 \right]^+ \\ &\stackrel{(a)}{\leq} \left[\left(\sum_{\{i \in \Omega: \rho_i < 1-L^{-1}\}} \rho_i \right) \frac{\log L + \log \frac{|\{i \in \Omega: \rho_i < 1-L^{-1}\}|}{\sum_{\{i \in \Omega: \rho_i < 1-L^{-1}\}} \rho_i}}{D_{min}} - K'_1 \right]^+ \\ &\stackrel{(b)}{\leq} \left[\frac{2 + L^{-1} \log(M-1)}{D_{min}} - K'_1 \right]^+, \end{aligned} \quad (55)$$

where (a) follows by Jensen's inequality; and (b) follows since $L > 1$, $\sum_{\{i \in \Omega: \rho_i < 1-L^{-1}\}} \rho_i < L^{-1}$ for any $\boldsymbol{\rho} \notin \mathbb{P}_L(\Theta)$, and $x \log \frac{1}{x} \leq 1$ for $x \in [0, 1]$. In other words, for $K'_1 \geq \frac{2+L^{-1} \log(M-1)}{D_{min}}$,

$$\underline{V}_1(\boldsymbol{\rho}) = 0 \leq \min_{j \in \Omega} (1 - \rho_j) L \quad \forall \boldsymbol{\rho} \notin \mathbb{P}_L(\Theta).$$

On the other hand, for all $\boldsymbol{\rho} \in \mathbb{P}_L(\Theta)$, we have

$$\begin{aligned} \underline{V}_1(\boldsymbol{\rho}) &\leq \left[\sum_{i=1}^M \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i}{1-\rho_i}}{D_{min}} - K'_1 \right]^+ \\ &= \left[\frac{\log(L-1)}{D_{min}} + \frac{\sum_{i=1}^M \rho_i \log \frac{1-\rho_i}{\rho_i}}{D_{min}} - K'_1 \right]^+. \end{aligned} \quad (56)$$

Now let

$$\begin{aligned} f_1(L, \boldsymbol{\rho}) &:= \frac{\log(L-1)}{D_{min}} + \frac{\sum_{i=1}^M \rho_i \log \frac{1-\rho_i}{\rho_i}}{D_{min}} \\ &\stackrel{(a)}{=} \frac{\log(L-1)}{D_{min}} + \frac{\sum_{i=1}^{M-1} \rho_i \log \frac{1-\rho_i}{\rho_i} + (1 - \sum_{i=1}^{M-1} \rho_i) \log \frac{(\sum_{i=1}^{M-1} \rho_i)}{(1 - \sum_{i=1}^{M-1} \rho_i)}}{D_{min}}, \end{aligned} \quad (57)$$

where (a) holds since $\sum_{i=1}^M \rho_i = 1$. Furthermore, let $\alpha = \max\{1, \frac{2}{D_{\min}}\}$, $\beta = \frac{1}{3}(D_{\min} - \frac{\log e}{\alpha})$, and $L^* = \max\{2\alpha, \frac{4}{\beta}, \frac{1}{\beta^2}, \frac{\log(M-1)}{\beta}\}$. Next we show that for all $L \geq L^*$ and at belief vector $\boldsymbol{\rho} = [\frac{\alpha}{(M-1)L}, \dots, \frac{\alpha}{(M-1)L}, 1 - \frac{\alpha}{L}]$, function $f_1(L, \cdot)$ has an upper bound independent of L while its partial derivatives with respect to ρ_i , $i = 1, 2, \dots, M-1$ are less than L :

$$\begin{aligned}
f_1(L, \boldsymbol{\rho}) \Big|_{\boldsymbol{\rho} = [\frac{\alpha}{(M-1)L}, \dots, \frac{\alpha}{(M-1)L}, 1 - \frac{\alpha}{L}]} &= \frac{\log(L-1)}{D_{\min}} + \frac{\sum_{i=1}^{M-1} \frac{\alpha}{(M-1)L} \log \frac{1 - \frac{\alpha}{(M-1)L}}{\frac{\alpha}{(M-1)L}} + (1 - \frac{\alpha}{L}) \log \frac{\frac{\alpha}{L}}{1 - \frac{\alpha}{L}}}{D_{\min}} \\
&= \frac{\log \frac{\alpha(L-1)}{L-\alpha}}{D_{\min}} + \frac{\alpha}{L} \frac{\log \frac{1 - \frac{\alpha}{(M-1)L}}{\frac{\alpha}{(M-1)L}} + \log \frac{1 - \frac{\alpha}{L}}{\frac{\alpha}{L}}}{D_{\min}} \\
&= \frac{\log \frac{\alpha(L-1)}{L-\alpha}}{D_{\min}} + \frac{\alpha}{L} \frac{\log(M-1) + 2 \log \frac{\alpha}{L}}{D_{\min}} \\
&\stackrel{(a)}{\leq} \frac{\log \frac{\alpha(L-1)}{L-\alpha}}{D_{\min}} + \frac{\alpha\beta + 2}{D_{\min}} \\
&= \frac{\log \frac{\alpha}{1 - \frac{\alpha-1}{L-1}}}{D_{\min}} + \frac{\alpha\beta + 2}{D_{\min}} \\
&\stackrel{(b)}{\leq} \frac{\log(2\alpha - 1) + \alpha\beta + 2}{D_{\min}}, \tag{58}
\end{aligned}$$

where (a) follows from the fact that $x \log \frac{1}{x} \leq 1$ for $x \in [0, 1]$ and $L \geq \frac{\log(M-1)}{\beta}$; and (b) holds since $L \geq 2\alpha$. And,

$$\begin{aligned}
\frac{\partial f_1}{\partial \rho_i}(L, \boldsymbol{\rho}) \Big|_{\boldsymbol{\rho} = [\frac{\alpha}{(M-1)L}, \dots, \frac{\alpha}{(M-1)L}, 1 - \frac{\alpha}{L}]} &= \left(\log \frac{1 - \rho_i}{\rho_i} - \frac{\log e}{1 - \rho_i} - \log \frac{1 - \rho_M}{\rho_M} + \frac{\log e}{1 - \rho_M} \right) \frac{1}{D_{\min}} \Big|_{\boldsymbol{\rho} = [\frac{\alpha}{(M-1)L}, \dots, \frac{\alpha}{(M-1)L}, 1 - \frac{\alpha}{L}]} \\
&= \left(\log \frac{(M-1)L - \alpha}{\alpha} - \frac{\log e}{1 - \frac{\alpha}{(M-1)L}} - \log \frac{\alpha}{L - \alpha} + \frac{\log e}{\frac{\alpha}{L}} \right) \frac{1}{D_{\min}} \\
&\leq \left(\log(M-1) + 2 \log L + \frac{\log e}{\alpha} L \right) \frac{1}{D_{\min}} \\
&= L + \left(\log(M-1) + 2 \log L - \left(D_{\min} - \frac{\log e}{\alpha} \right) L \right) \frac{1}{D_{\min}} \\
&\leq L + (\log(M-1) + 2 \log L - 3\beta L) \frac{1}{D_{\min}} \\
&\leq L + (\log L - \beta L) \frac{2}{D_{\min}} \\
&\stackrel{(a)}{\leq} L + (\log L - \log(\beta L)^2) \frac{2}{D_{\min}} \\
&\leq L, \tag{59}
\end{aligned}$$

where (a) follows from the fact that $\log x^2 \leq x$ for $x \geq 4$.

From symmetry and concavity of $f_1(L, \cdot)$, (58), (59), it is clear that for $L \geq L^*$, and for all $\boldsymbol{\rho} \in \mathbb{P}_L(\Theta)$,

$$\left[f_1(L, \boldsymbol{\rho}) - \frac{\log(2\alpha - 1) + \alpha\beta + 2}{D_{\min}} \right]^+ \leq \min_{j \in \Omega} (1 - \rho_j)L. \quad (60)$$

Fig. 3 shows this for $M = 2$. Furthermore, for all $L < L^*$,

$$f_1(L, \boldsymbol{\rho}) \leq f_1(L^*, \boldsymbol{\rho}) \leq \max_{\hat{\boldsymbol{\rho}}} f_1(L^*, \hat{\boldsymbol{\rho}}) \leq \frac{\log L^* + \log M}{D_{\min}}.$$

This together with (55), (56), and (60) implies the assertion of the claim for

$$K'_1 = \frac{1 + \alpha\beta + \log L^* + \log M}{D_{\min}} \geq \frac{\max\{2 + L^{-1} \log(M - 1), \log L^* + \log M, \log(2\alpha - 1) + \alpha\beta + 2\}}{D_{\min}}.$$

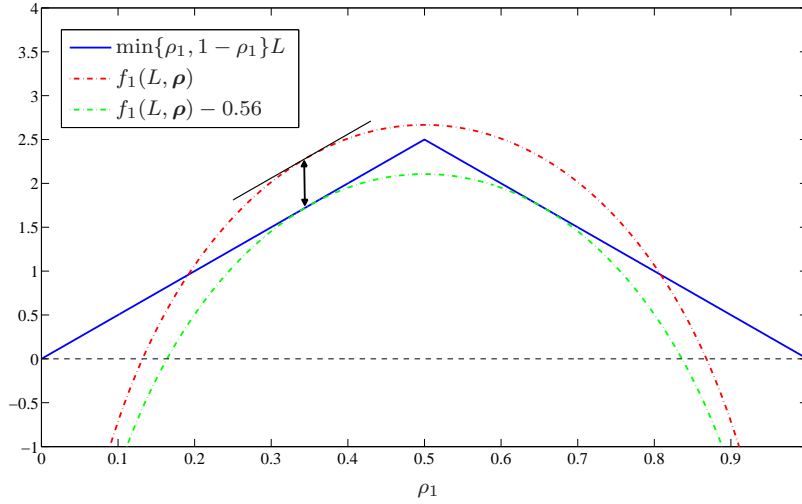


Fig. 3. Computing K'_1 for $M = 2$, $L = 5$, and $D_{\min} = .75$. In this example, the derivative of f_1 with respect to ρ_1 is equal to L at $\rho_1 = 0.35$ and $K'_1 \geq 0.56$ ensures that $f_1(L, \boldsymbol{\rho}) - K'_1 \leq \min\{\rho_1, 1 - \rho_1\}L$. We have $\alpha = 2.67$, $\beta = 0.07$, $L^* = 206.06$, and $K'_1 = 13.16$.

B. Proof of Claim 2

For any $\boldsymbol{\rho}$ such that $J'(\boldsymbol{\rho}) = 0$, the inequality holds trivially. For any $\boldsymbol{\rho}$ such that $J'(\boldsymbol{\rho}) > 0$ and for any action $a \in \mathcal{A}$ we have

$$\begin{aligned} (\mathbb{T}^a J')(\boldsymbol{\rho}) &\geq \sum_{i=1}^M \int \rho_i q_i^a(z) \frac{\log \frac{1-L^{-1}}{L^{-1}} + \log \xi - \log \frac{\rho_i q_i^a(z)}{\sum_{k \neq i} \rho_k q_k^a(z)}}{D_{\max}} dz - K'_2 \\ &= J'(\boldsymbol{\rho}) - \sum_{i=1}^M \rho_i \frac{\int q_i^a(z) \log \frac{q_i^a(z)}{\sum_{k \neq i} \frac{\rho_k}{1-\rho_i} q_k^a(z)} dz}{D_{\max}} \\ &\stackrel{(a)}{\geq} J'(\boldsymbol{\rho}) - \sum_{i=1}^M \rho_i \frac{\sum_{k \neq i} \frac{\rho_k}{1-\rho_i} D(q_i^a \| q_k^a)}{D_{\max}} \\ &\geq J'(\boldsymbol{\rho}) - 1, \end{aligned}$$

where (a) follows from Jensen's inequality.

C. Proof of Claim 3

For all $\boldsymbol{\rho}$ satisfying $\max_{i \in \Omega} \rho_i > 1 - \delta$,

$$\begin{aligned} H(\boldsymbol{\rho}) &< (1 - \delta) \log \frac{1}{1 - \delta} + (M - 1) \times \frac{\delta}{M - 1} \log \frac{1}{\delta/(M - 1)} \\ &= H([\delta, 1 - \delta]) + \delta \log(M - 1), \end{aligned}$$

and hence, $J'' < 0$. In other words, $J(\boldsymbol{\rho}) = J''(\boldsymbol{\rho}) > 0$ implies that $\max_{i \in \Omega} \rho_i \leq 1 - \delta$.

Let $\hat{\boldsymbol{\rho}} = \Phi^a(\boldsymbol{\rho}, z)$. Inequality (24) holds trivially if $\max_{i \in \Omega} \hat{\rho}_i \leq 1 - \delta$ since $J(\hat{\boldsymbol{\rho}}) \geq J''(\hat{\boldsymbol{\rho}})$ and $J''(\hat{\boldsymbol{\rho}})$ is greater than or equal to the right-hand side of (24). If $\max_{i \in \Omega} \hat{\rho}_i > 1 - \delta$, we get

$$\begin{aligned} J(\hat{\boldsymbol{\rho}}) &= J'(\hat{\boldsymbol{\rho}}) \\ &= \left[\sum_{i=1}^M \hat{\rho}_i \frac{\log \frac{1-L^{-1}}{L^{-1}} + \log \xi - \log \frac{\hat{\rho}_i}{1-\hat{\rho}_i}}{D_{max}} - K'_2 \right]^+ \\ &\stackrel{(a)}{\geq} \left[\sum_{i=1}^M \hat{\rho}_i \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} - K'_2 \right]^+ \\ &\geq \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} - K'_2, \end{aligned}$$

where (a) follows from the fact that under Assumption 2 and for all $i \in \Omega$,

$$\begin{aligned} \log \frac{\hat{\rho}_i}{1 - \hat{\rho}_i} &\leq \left| \log \frac{\hat{\rho}_i}{1 - \hat{\rho}_i} - \log \frac{\rho_i}{1 - \rho_i} \right| + \left| \log \frac{\rho_i}{1 - \rho_i} \right| \\ &\leq \left| \log \frac{\rho_i q_i^a(z)}{\sum_{j \neq i} \rho_j q_j^a(z)} - \log \frac{\rho_i}{1 - \rho_i} \right| + \log \frac{1 - \delta}{\delta} \\ &= \left| \log \frac{q_i^a(z)}{\sum_{j \neq i} \frac{\rho_j}{1 - \rho_j} q_j^a(z)} \right| + \log \frac{1 - \delta}{\delta} \\ &\leq \log \xi + \log \frac{1 - \delta}{\delta}. \end{aligned}$$

D. Proof of Claim 4

Following similar lines as the proof of Claim 1, we can select K'_2 sufficiently large such that $J'(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j)L$. Recall that $J'(\boldsymbol{\rho}) = [f_2(L, \boldsymbol{\rho}) - K'_2]^+$, where

$$f_2(L, \boldsymbol{\rho}) := \frac{\log(L - 1) + \log \xi}{D_{max}} + \frac{\sum_{i=1}^M \rho_i \log \frac{1 - \rho_i}{\rho_i}}{D_{max}}. \quad (61)$$

By the assumption of Claim 4, there exists $\kappa > 0$ such that $\log M \leq (I_{\max} - \kappa)L$ for all M . Set $\alpha = \max\{1, \frac{2}{\kappa}\}$, $\beta = \frac{1}{2}(\kappa - \frac{\log e}{\alpha})$, and let $L'_2 = \max\{2\alpha, \frac{4}{\beta}, \frac{1}{\beta^2}\}$. For $L \geq L'_2$ and for all $i = 1, 2, \dots, M-1$, we obtain

$$\begin{aligned} f_2(L, \boldsymbol{\rho}) \Big|_{\boldsymbol{\rho} = [\frac{\alpha}{(M-1)L}, \dots, \frac{\alpha}{(M-1)L}, 1 - \frac{\alpha}{L}]} &= \frac{\log \frac{\alpha(L-1)}{L-\alpha} + \log \xi}{D_{\max}} + \frac{\alpha \log(M-1) + 2 \log \frac{\alpha}{L}}{L D_{\max}} \\ &\leq \frac{\log \frac{\alpha(L-1)}{L-\alpha} + \log \xi}{D_{\max}} + \frac{\alpha(I_{\max} - \kappa) + 2}{D_{\max}} \\ &\leq \frac{\log(2\alpha - 1) + \log \xi + \alpha(I_{\max} - \kappa) + 2}{D_{\max}}, \end{aligned} \quad (62)$$

and,

$$\begin{aligned} \frac{\partial f_2}{\partial \rho_i}(L, \boldsymbol{\rho}) \Big|_{\boldsymbol{\rho} = [\frac{\alpha}{(M-1)L}, \dots, \frac{\alpha}{(M-1)L}, 1 - \frac{\alpha}{L}]} &\leq \left(\log(M-1) + 2 \log L + \frac{\log e}{\alpha} L \right) \frac{1}{D_{\max}} \\ &\leq \left(2 \log L + (I_{\max} - \kappa + \frac{\log e}{\alpha}) L \right) \frac{1}{D_{\max}} \\ &\leq L + (\log L - \beta L) \frac{2}{D_{\max}} \\ &\leq L. \end{aligned} \quad (63)$$

From symmetry and concavity of $f_2(L, \cdot)$, (62), (63), it is clear that for $L \geq L'_2$, and for all $\boldsymbol{\rho} \in \mathbb{P}_L(\Theta)$,

$$\left[f_2(L, \boldsymbol{\rho}) - \frac{\log(2\alpha - 1) + \log \xi + \alpha(I_{\max} - \kappa) + 2}{D_{\max}} \right]^+ \leq \min_{j \in \Omega} (1 - \rho_j) L. \quad (64)$$

Furthermore, for $L < L'_2$, we have

$$f_2(L, \boldsymbol{\rho}) \leq f_2(L'_2, \boldsymbol{\rho}) \leq \max_{\hat{\boldsymbol{\rho}}} f_2(L'_2, \hat{\boldsymbol{\rho}}) \leq \frac{\log L'_2 + \log \xi + \log M}{D_{\max}} \leq \frac{\log L'_2 + \log \xi + L'_2(I_{\max} - \kappa)}{D_{\max}}.$$

In other words, selecting

$$K'_2 \geq \frac{\max\{\log L'_2 + \log \xi + L'_2(I_{\max} - \kappa), \log(2\alpha - 1) + \log \xi + \alpha(I_{\max} - \kappa) + 2\}}{D_{\max}}, \quad (65)$$

satisfies $J'(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j) L$.

Next we discuss on the selection of K'_2 such that $J''(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j) L$ is also satisfied. Let

$$f_3(L, \boldsymbol{\rho}) := \frac{H(\boldsymbol{\rho}) - H([\delta, 1 - \delta]) - \delta \log(M-1)}{I_{\max}},$$

and rewrite,

$$J''(\boldsymbol{\rho}) = \left[f_3(L, \boldsymbol{\rho}) + \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{\max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - K'_2 \right]^+.$$

We show that at belief vector $\boldsymbol{\rho} = [\frac{0.5}{(M-1)L}, \dots, \frac{0.5}{(M-1)L}, 1 - \frac{0.5}{L}]$, and for $L \geq L_2'' := \max\{\frac{4I_{max}}{\kappa}, \frac{I_{max}^2}{\kappa^2}\}$, $f_3(L, \boldsymbol{\rho})$ has an upper bound independent of L and its partial derivatives with respect to ρ_i , $i = 1, 2, \dots, M-1$ are less than L . In other words,

$$\begin{aligned} & f_3(L, \boldsymbol{\rho}) \Big|_{\boldsymbol{\rho} = [\frac{0.5}{(M-1)L}, \dots, \frac{0.5}{(M-1)L}, 1 - \frac{0.5}{L}]} \\ &= \frac{H([\frac{1}{2L}, 1 - \frac{1}{2L}]) + \frac{1}{2L} \log(M-1) - H([\delta, 1 - \delta]) - \delta \log(M-1)}{I_{max}} \\ &\leq \frac{1 + \frac{1}{2}(I_{max} - \kappa)}{I_{max}}, \end{aligned}$$

and,

$$\begin{aligned} \frac{\partial f_3}{\partial \rho_i}(L, \boldsymbol{\rho}) \Big|_{\boldsymbol{\rho} = [\frac{0.5}{(M-1)L}, \dots, \frac{0.5}{(M-1)L}, 1 - \frac{0.5}{L}]} &= \frac{1}{I_{max}} \log \frac{\rho_M}{\rho_i} \Big|_{\boldsymbol{\rho} = [\frac{0.5}{(M-1)L}, \dots, \frac{0.5}{(M-1)L}, 1 - \frac{0.5}{L}]} \\ &= \frac{1}{I_{max}} \log \frac{1 - \frac{0.5}{L}}{\frac{0.5}{(M-1)L}} \\ &\leq \frac{1}{I_{max}} (\log(M-1) + \log L) \\ &\leq \frac{1}{I_{max}} (L(I_{max} - \kappa) + \log L) \\ &= L - \left(\frac{\kappa}{I_{max}} L - \log L \right) \\ &\leq L. \end{aligned}$$

Furthermore, for $L < L_2''$, we have

$$f_3(L, \boldsymbol{\rho}) \leq f_3(L_2'', \boldsymbol{\rho}) \leq \max_{\hat{\boldsymbol{\rho}}} f_3(L_2'', \boldsymbol{\rho}) \leq \frac{\log M}{I_{max}} \leq \frac{L_2''(I_{max} - \kappa)}{I_{max}}.$$

We also note that for all K_2' satisfying (65),

$$\frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - K_2' \leq J'(\boldsymbol{\rho}) \leq \min_{j \in \Omega} (1 - \rho_j)L.$$

Thus, if K_2' satisfies (65) as well as the following condition,

$$K_2' \geq \frac{\max\{L_2''(I_{max} - \kappa), 1 + \frac{1}{2}(I_{max} - \kappa)\}}{I_{max}}, \quad (66)$$

we have

$$\left[f_3(L, \boldsymbol{\rho}) + \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{1-\delta}{\delta}}{D_{max}} \mathbf{1}_{\{\max_{i \in \Omega} \rho_i \leq 1-\delta\}} - K_2' \right]^+ \leq \min_{j \in \Omega} (1 - \rho_j)L.$$

By selecting K_2' to be independent from L and M and larger than (65) and (66), we have the assertion of the claim. The following selection of K_2' satisfies the above conditions:

$$K_2' = \max \left\{ \frac{\log L_2' + \log \xi + L_2'(\bar{I}_{max} - \kappa) + 2}{\underline{D}_{max}}, \frac{4\bar{I}_{max}}{\kappa} + \frac{\bar{I}_{max}^2}{\kappa^2} + \frac{1}{\underline{I}_{max}} \right\}. \quad (67)$$

E. Proof of Claim 5

Let $D_\lambda = \min_{i,j \in \Omega} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a || q_j^a)$. Here we bound \underline{V}_λ as follows:

$$\begin{aligned} \underline{V}_\lambda(\rho) &\leq \left[\frac{2 + L^{-1} \log(M-1)}{D_\lambda} - K'_\lambda \right]^+, \quad \forall \rho \notin \mathbb{P}_L(\Theta) \\ \underline{V}_\lambda(\rho) &\leq \left[\sum_{i=1}^M \rho_i \frac{\log \frac{1-L^{-1}}{L^{-1}} - \log \frac{\rho_i}{1-\rho_i}}{D_\lambda} - K'_\lambda \right]^+, \quad \forall \rho \in \mathbb{P}_L(\Theta). \end{aligned}$$

Following similar lines as the proof of Claim 1, we can show that

$$K'_\lambda = \frac{1 + \alpha\beta + \log L^* + \log M}{D_\lambda} \geq \frac{\max\{2 + L^{-1} \log(M-1), \log L^* + \log M, \log(2\alpha - 1) + \alpha\beta + 2\}}{D_\lambda}$$

ensures that $\underline{V}_\lambda(\rho) \leq \min_{j \in \Omega} (1 - \rho_j)L$ where $\alpha = \max\{1, \frac{2}{D_\lambda}\}$, $\beta = \frac{1}{3} (D_\lambda - \frac{\log e}{\alpha})$, and $L^* = \max\{2\alpha, \frac{4}{\beta}, \frac{1}{\beta^2}, \frac{\log(M-1)}{\beta}\}$.

ACKNOWLEDGEMENTS

We would like to thank Sergio Verdú, Young-Han Kim, Todd Coleman, Maxim Raginsky, Michèle Wigger, and Yury Polyanskiy for valuable discussions and suggestions.

REFERENCES

- [1] A. Wald and J. Wolfowitz, “Optimal character of the sequential probability ratio tests,” *The Annals of Mathematical Statistics*, vol. 19, no. 3, pp. 326–339, 1948.
- [2] P. Armitage, “Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis,” *Journal of the Royal Statistical Society, Series B*, vol. 12, no. 1, pp. 137–144, 1950.
- [3] G. Lorden, “Nearly-optimal sequential tests for finitely many parameter values,” *The Annals of Statistics*, vol. 5, no. 1, pp. 1–21, 1977.
- [4] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, “Multihypothesis sequential probability ratio tests. I. Asymptotic optimality,” *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2448–2461, November 1999.
- [5] H. Chernoff, “Sequential design of experiments,” *The Annals of Mathematical Statistics*, vol. 30, pp. 755–770, 1959.
- [6] S. M. Berry, B. P. Carlin, J. J. Lee, and P. Müller, *Bayesian adaptive methods for clinical trials*. CRC Press, 2010.
- [7] P. Shenoy and A. J. Yu, “Rational decision-making in inhibitory control,” *Frontiers in Human Neuroscience*, vol. 5, no. 48, 2011.
- [8] A. O. Hero and D. Cochran, “Sensor management: past, present, and future,” *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3064–3075, December 2011.
- [9] G. A. Hollinger, U. Mitra, and G. S. Sukhatme, “Active classification: theory and application to underwater inspection,” 2011, available on arXiv:1106.5829.
- [10] R. D. Nowak, “The geometry of generalized binary search,” *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7893–7906, December 2011.
- [11] M. V. Burnashev, “Data transmission over a discrete channel with feedback. Random transmission time,” *Problemy Peredachi Informatsii*, vol. 12, no. 4, pp. 10–30, 1975.
- [12] D. Blackwell, “Equivalent comparisons of experiments,” *The Annals of Mathematical Statistics*, vol. 24, pp. 265–272, 1953.

- [13] D. V. Lindley, "On a measure of the information provided by an experiment," *The Annals of Mathematical Statistics*, vol. 27, no. 4, pp. 986–1005, 1956.
- [14] L. L. Cam, "Sufficiency and approximate sufficiency," *The Annals of Mathematical Statistics*, vol. 35, no. 4, pp. 1419–1455, 1964.
- [15] M. H. DeGroot, *Optimal statistical decisions*. McGraw-Hill, Inc., 1970.
- [16] K. Goel and M. H. DeGroot, "Comparison of experiments and information measures," *The Annals of Statistics*, vol. 7, no. 5, pp. 1066–1077, 1979.
- [17] E. L. Lehmann, "Comparing location experiments," *The Annals of Statistics*, vol. 16, no. 2, pp. 521–533, 1988.
- [18] E. Torgersen, *Stochastic orders and comparison of experiments*. Hayward, CA: Institute of Mathematical Statistics, 1991, vol. 19, pp. 334–371.
- [19] M. Naghshvar and T. Javidi, "Performance bounds for active sequential hypothesis testing," in *IEEE International Symposium on Information Theory (ISIT)*, 2011, pp. 2666–2670.
- [20] M. H. DeGroot, "Uncertainty, information, and sequential experiments," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. 404–419, 1962.
- [21] M. Naghshvar and T. Javidi, "Active hypothesis testing: sequentiality and adaptivity gains," in *Conference on Information Sciences and Systems (CISS)*, March 2012.
- [22] W. J. Blot and D. A. Meeter, "Sequential experimental design procedures," *Journal of the American Statistical Association*, vol. 68, no. 343, pp. 586–593, 1973.
- [23] S. Nitinawarat, G. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," 2012, available on arXiv:1205.0858.
- [24] A. E. Albert, "The sequential design of experiments for infinitely many states of nature," *The Annals of Mathematical Statistics*, vol. 32, pp. 774–799, 1961.
- [25] J. Kiefer and J. Sacks, "Asymptotically optimum sequential inference and design," *The Annals of Mathematical Statistics*, vol. 34, no. 3, pp. 705–750, 1963.
- [26] R. Keener, "Second order efficiency in the sequential design of experiments," *The Annals of Statistics*, vol. 12, no. 2, pp. 510–532, 1984.
- [27] S. P. Lalley and G. Lorden, "A control problem arising in the sequential design of experiments," *The Annals of Probability*, vol. 14, no. 1, pp. 136–172, 1986.
- [28] S. Nitinawarat, G. Atia, and V. V. Veeravalli, "Controlled sensing for hypothesis testing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2012.
- [29] G. Atia and V. V. Veeravalli, "Controlled sensing for sequential multihypothesis testing," in *IEEE International Symposium on Information Theory (ISIT)*, July 2012.
- [30] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, pp. 8–19, 1956.
- [31] T. M. Cover and J. A. Thomas, *Elements of information theory* (2. ed.). John Wiley & Sons, Inc., 2006.
- [32] R. G. Gallager, *Information theory and reliable communication*. John Wiley & Sons, Inc., New York, 1968.
- [33] G. Lorden, "Asymptotic efficiency of three-stage hypothesis tests," *The Annals of Statistics*, vol. 11, pp. 129–140, 1983.
- [34] J. Bartroff, "Asymptotically optimal multistage tests of simple hypotheses," *The Annals of Statistics*, vol. 35, no. 5, pp. 2075–2105, 2007.
- [35] P. R. Kumar and P. Varaiya, *Stochastic systems: estimation, identification, and adaptive control*. Prentice-Hall, Inc., 1986.
- [36] D. P. Bertsekas and S. E. Shreve, *Stochastic optimal control: the discrete-time case*. Athena Scientific, 2007.
- [37] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, Inc., 1994.
- [38] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, August 2011.
- [39] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 405–417, July 1974.

- [40] I. Csiszár and P. Shields, “Information theory and statistics: a tutorial,” *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 4, pp. 417–528, December 2004.
- [41] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, “Reliability criteria in information theory and in statistical hypothesis testing,” *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 2, pp. 97–263, January 2007.
- [42] M. Hayashi, “Discrimination of two channels by adaptive methods and its application to quantum system,” *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3807–3820, August 2009.
- [43] Y. Polyanskiy and S. Verdú, “Hypothesis testing with feedback,” in *Information Theory and Applications Workshop (ITA)*, 2011.
- [44] D. A. Castanon, “Optimal search strategies in dynamic hypothesis testing,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 7, pp. 1130–1138, July 1995.
- [45] M. Horstein, “Sequential transmission using noiseless feedback,” *IEEE Transactions on Information Theory*, vol. 9, no. 3, pp. 136–143, July 1963.
- [46] M. V. Burnashev and K. S. Zigangirov, “An interval estimation problem for controlled observations,” *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 51–61, 1974.
- [47] M. Naghshvar and T. Javidi, “Variable-length coding with noiseless feedback and finite messages,” in *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, 2010, pp. 317–321.
- [48] A. Mahajan, A. Nayyar, and D. Teneketzis, “Identifying tractable decentralized problems on the basis of information structures,” in *Proceedings of the 46th Allerton conference on communication, control, and computing*, 2008, pp. 1440–1449.
- [49] P. Berlin, B. Nakiboglu, B. Rimoldi, and E. Telatar, “A simple converse of Burnashev’s reliability function,” *IEEE Transactions on Information Theory*, vol. 55, pp. 3074–3080, 2009.
- [50] M. Naghshvar and T. Javidi, “Active M -ary sequential hypothesis testing,” in *IEEE International Symposium on Information Theory (ISIT)*, 2010, pp. 1623–1627.
- [51] C. McDiarmid, “On the method of bounded differences,” *Surveys in Combinatorics, London Mathematical Society Lecture Note Series 141*, Cambridge University Press, pp. 148–188, 1989.

Mohammad Naghshvar (S’07) received the B.S. degree in electrical engineering from Sharif University of Technology in 2007. He obtained his M.S. degree in 2009 in electrical and computer engineering from University of California San Diego where he is currently pursuing his studies toward a Ph.D. degree. His research interests include active hypothesis testing and optimal experimental design, stochastic control and optimization, wireless communication and information theory, routing and scheduling in wireless networks.

Tara Javidi (S’96-M’02) studied electrical engineering at the Sharif University of Technology from 1992 to 1996. She received her MS degrees in Electrical Engineering: Systems, and Applied Mathematics: Stochastics, from the University of Michigan, Ann Arbor, MI. She received her PhD in electrical engineering and computer science from the University of Michigan, Ann Arbor, in 2002. From 2002 to 2004, she was an assistant professor at the Electrical Engineering Department, University of Washington, Seattle. In 2005, she joined University of California, San Diego, where she is currently an associate professor of electrical and computer engineering.

Tara Javidi was a Barbour Scholar during 1999-2000 academic year and received an NSF CAREER Award in 2004. Her research interests are in communication networks, stochastic resource allocation, stochastic control theory, and wireless communications.